The Journal of Applied Psychology

Vol. VI MARCH, 1922 No. 1

EDITORS

JAMES P. PORTER
Clark University

WILLIAM F. BOOK
Indiana University

And a Board of Co-operating Editors

CLARK UNIVERSITY
WORCESTER MASS.
FLORENCE CHANDLER, Publisher

Entered as second class matter June 9, 1917, at the post office at Worcester, Mass. under the Act of March 3, 1879.

Acceptance for mailing at special rate of postage provided for in section 1103, Act of October 3, 1917, Authorized April 20, 1921.

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. Bingham, Carnegie Institute of Technology; June E. Downey, University of Wyoming; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; Joseph Jastrow, University of Wisconsin; Herbert S. Langfeld, Harvard University; James H. Leuba, Bryn Mawr College; Henry C. McComas, Princeton University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, University of Wisconsin; Edward K. Strong, Jr., Carnegie Institute of Technology; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; J. E. Wallace Wallin, Psycho-Educational Clinic, St. Louis; G. M. Whipple, University of Michigan; Helen Thompson Woolley, Vocation Bureau, Cincinnati.

The scope of the Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environmental conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

Volume VI — March, 1922 — Number 1

TABLE OF CONTENTS

Rose C. Anderson. Methods and Results of Mental Surveys 1

Edward L. Thorndike. On Finding Equivalent Scores in Tests of Intelligence 29

Anna Ross Reyburn. Some Factors of Æsthetic Judgment 34

Harry Dexter Kitson. Minor Studies in the Psychology of Advertising from Psychological Laboratory of Indiana University...

V. The Development of Art-Forms in Magazine Advertising. 59

VI. Color in Magazine Advertising 64

VII. Negative Suggestion in Advertising 66

M. J. Ream. A Social Relations Test 69

A. T. Poffenberger and V. H. Vartanian. The Letter of Application in Vocational Selection 74

News and Comment 81

Book Reviews 88

Communications, manuscripts, reviews, summaries, notices, books and off-prints for review should be addressed to Dr. James P. Porter, Clark University, Worcester, Mass. Contributors are requested to follow the conventional method of citation employed by this Journal; in every case year, volume and pages should be specified.

The Journal of Applied Psychology is published quarterly. The subscription price is $4.00 a year. Single numbers $1.25. Remittances and business communications should be addressed to Florence Chandler, Publisher, Clark University, (950 Main St.), Worcester, Mass.

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. VI MARCH, 1922 No. 1

METHODS AND RESULTS OF MENTAL SURVEYS

By ROSE G. ANDERSON, Mental Examiner, Research Bureau, Minnesota State Board of Control

Studies in feeble-mindedness during the past few years have once more brought to the foreground the old questions: who are the feeble-minded and how many are there? The chief method of arriving at answers to these questions, particularly the latter, has been through the making of mental surveys. The first important survey of this kind was made as recently as 1904 by the British Royal Commission in its well-known study of typical districts of the British Isles.[1] This survey consisted of an enumeration of "all mentally defective persons" in these districts. The method adopted consisted of visits to all public elementary schools, and to all institutions where there would be any likelihood of finding any cases of mental defect. In addition, physicians, clergy, police, charitable organizations, and any other possible sources of information were consulted. The frequency of mental deficiency in the sixteen areas surveyed was found to be 3.28 per thousand inhabitants or .328 of one per cent.

Subsequent to the survey of the British Royal Commission, a number of others have been conducted in various parts of this country. Whether these surveys were inspired by the British survey or not, many of them have been conducted in much the same manner. In recent years some mental testing has been done in selected cases, and, in one instance at least, both group and individual tests have been given in the schools surveyed. In the case of all the surveys some conclusion has been reached in regard to the per cent of feeble-minded in the school population or the general population. The extent to which these estimates vary would perhaps be a little disconcert-

[1] Tredgold, A. F., Mental Deficiency, Bailliere, Tindall and Cox, London, 1920.

ing to one trying to find the general conclusion reached in these surveys. The one point on which there seems to be agreement is that the estimate finally made is the lowest that could be made, or that if it errs, the error is in the direction of being too conservative, rather than as liberal as the facts warranted.

It is the object of this paper: first, to review briefly some of these surveys, summarizing their methods and results; second, to point out the variation in the results, and to attempt to show that the lack of agreement is due mostly to the differences in the methods and purposes of the various surveys; and third, to present in some detail the methods and results of two types of surveys made by the Department of Research of the Minnesota School for the Feeble-Minded under the direction of Dr. F. Kuhlmann.[2]

Summary of Methods and Findings of Surveys

In the survey of a rural county of Ohio, the purpose was to discover the extent and social significance of feeble-mindedness in that particular part of the state.[3] Information was secured from the public schools, physicians, township trustees, county institutions, county and city officials, etc. The definition of the English Royal Commission was adopted as the standard, except in the case of school children, so that the criterion was principally a social one. The plan followed in detecting the feeble-minded in the schools was to select those who were three or more years retarded in their grades, together with others suggested by the teacher. These were then considered individually from a number of different angles, with the question always in mind whether they would be able to function normally when mature. Summarizing the results briefly, the percent of feeble-minded in the school population was found to be 1.8, and in the general population, 1.06.[4] The incidence of feeble-mindedness for rural communities was more than two and one-half times what it was for urban communities.

Two reports are available of surveys of New Castle County,

[2] This Department was moved to Saint Paul, Sept. 1, 1921. It has been reorganized into the Research Bureau and is now a separate state department under the State Board of Control.

[3] Sessions, Mina A., The Feeble-Minded in a Rural County of Ohio, Bulletin No. 6 of the Bureau of Juvenile Research. Publication No. 12, Ohio Board of Administration, 1918.

[4] See Miner, Deficiency and Delinquency, Warwick & York, Baltimore, 1918, pp. 48-50, on discrepancy between estimates for school population and general population.

Delaware, one conducted by the Children's Bureau,[5] and one by the United States Public Health Service.[6] The former was primarily a social study to determine conditions under which mental defectives live, problems involved in the lack of proper facilities for their care, and the extent of the need for public supervision and institutional provision. A list of supposed mental defectives was secured through the cooperation of the usual institutions, officials, and social agencies. Mental examinations were obtained when possible. The examinations made by the United States Public Health Service of all school children outside of Wilmington, and of a selected list in Wilmington were made available. These were supplemented by diagnoses by "other physicians competent to determine mentality." In the case of inmates of institutions for dependents, only a few who had been previously examined or who were unquestionably feeble-minded, were included as positive cases. Out of 1100 cases reported to the investigators as possibly feeble-minded, 212 were classified as positive cases and an additional 361 as questionable. The latter group included those who were not given mental examinations and those for whom positive diagnoses could have been made only after more prolonged observation than it was possible for the examining physician to make. Miss Lundberg, however, concluded that adequate mental and social examinations would undoubtedly reveal a large number of positive cases among this latter group. The positive cases constitute a total of .16 of one percent of the population and the questionable cases .27 of one percent, making a total of .44 of one percent out of .84 of one percent reported as possible cases. In a table showing the mental defectives in school according to age, color, and grade attained, those least retarded pedagogically are the following: Grade I—8 yrs.; Grade II—10 yrs.; Grade III—12 yrs.; Grade IV—13 yrs.; Grade V—12 yrs.; Grade VI—16 yrs.

The survey of New Castle County by the United States Public Health Service was made in connection with a sanitary survey of the rural schools of the county, which also included the mental status of the children. Out of 3,793 children studied, .5 percent were classified as definitely feeble-minded and an additional 1.3 percent as probable mental defectives.

[5] Lundberg, Emma O., A Social Study of Mental Defectives in New Castle County, Delaware, Children's Bureau, United States Department of Labor, Bureau Publication No. 24, Government Printing Office, Washington, 1917.

[6] Mullan, E. H., Mental Status of Rural School Children, United States Public Health Service, Reprint No. 377 from the Public Health Reports, Government Printing Office, Washington, 1916.

Preliminary abbreviated mental tests and questions "suited in a general way to the child's age and school-grade" were used to select cases likely to be defective. These were further examined with the Goddard Revision of the Binet-Simon tests together with questions on general information, and other tests intended to throw light upon their "ability to see into, catch on, retain, improve," and a long list of other qualities. In a report of the test results of 209 "normal" children, Dr. Mullan included a chronological 15 testing 10.6, a chronological 14 testing 9.8, a chronological 13.5 testing 10.2, etc.

The United States Public Health Service also conducted a survey of the mental status of the rural school children of Porter County, Indiana.[7] A total of 2,185 children was examined. Special reliance was placed on the Binet test with modifications. In the case of exceptionally retarded children special methods were employed to determine the degree of mental impairment. These special methods are not described. In rating the children, the following were classified as exceptionally retarded: those of six years who were four mentally, of seven who were five, of eight who were five or six, of nine who were six or seven, of ten who were six or seven, of eleven who were seven or eight, of 12 or 13 who were eight or nine, of 14 or 15 who were nine or ten, of 16 who were 10, and of 17-20 who were 11. This group comprised all the children who were mentally deficient. Of the total school population, .9 percent were mental defectives.

The Maine survey by the National Committee for Mental Hygiene which was conducted by Dr. Guy Fernald was preliminary and educative in its nature and purposes, and did not attempt to discover all the cases of feeble-mindedness.[8] Information in regard to cases was secured from town officers, and in some larger towns and all cities conferences were arranged where a physician or a teacher was usually present. In many towns the layman's application of a definition given by the expert furnished the estimate of numbers. Estimates which were made by local observers "of unusual intelligence, experience, and opportunity for observation," were received from 121 institutions, including almshouses and jails. In a limited number of school rooms the Wells group tests of intelli-

[7] Clark, Taliaferro, The Mental Status of Rural School Children of Porter County, Indiana, United States Public Health Service, Reprinted from Public Health Bulletin No. 77, Government Printing Office, Washington, 1916.

[8] Report of the Maine Commission for the Feebleminded and of the Survey by the National Committee for Mental Hygiene, Augusta, Maine, 1918.

gence were given. A small number of cases (162) were intensively studied. In the individual "personality" examination, psychiatrical tests, the Stanford Revision of the Binet-Simon tests, and "researches in the field of character by the inductive method" were used. The study revealed a total of 1659 cases of feeble-mindedness, constituting .212 of one percent of the population of the state.

A number of surveys made in California have been published under the heading of *Surveys in Mental Deviation*.[9] Among these is the report of a study of the "Backward and Feeble-Minded Children in the Public Schools of 'X' County, California," planned by Dr. Terman and carried out under his direction. From teachers' ratings of their pupils into five classes on the basis of the quality of school-work done in the grade in which each child was at that time, and estimates of each pupil's mental ability, the lowest 20 percent in ability were selected. The plan was to test all suspected cases in all rural schools of the county and in "Y" city. The rule followed was to test every child rated by the teacher as seriously below average in school work or in native ability, and to test all those who were seriously over-age for their grade even if they did not fall in the former group. Twelve percent of these children were tested with the Stanford Revision of the Binet-Simon Intelligence Scale, a total of 174 out of 1464. The percent of feeble-mindedness for the rural schools was 4.9 and for "Y" city was 3.5 percent, the average for both being 4.24 percent in addition to which 1.98 percent were classed as borderzone cases. No child was placed below the borderzone group unless feeble-minded beyond all dispute. These percents take no account of the feeble-minded who had been kept at home, of those in the Sonoma State School, or of epileptics.

The survey of "X" County in Southern Indiana, under the direction of Drs. W. F. Book and U. G. Weatherly of Indiana University is an example of another method of procedure.[10]

---

[9] Surveys in Mental Deviation in Prisons, Public Schools, and Orphanages in California, California State Board of Charities and Corrections, California State Printing Office, Sacramento, 1918.

[10] Book, W. F., Variations in Mental Ability and Its Distribution among the School Population of an Indiana County, Bulletin of the Extension Division, Vol. IV, No. 4, Indianapolis, 1919.

Mental Defectives in Indiana, Second Report of the Indiana Committee on Mental Defectives, A Survey of Ten Counties, Indianapolis, 1919.

Pressey, S. L., and L. W., A Group Test Scale for Measuring General Intelligence, with First Results from 1,100 School Children, *Journal of Applied Psychology*, Sept., 1918, Vol. II, pp. 250-269.

The purpose of this survey was to determine the number of mental defectives of institutional grade in the county. A study was also made of children of superior intelligence. S. L. Pressey and L. W. Pressey had charge of the mental testing done in this survey and devised a group scale of intelligence which was given to all school children in grades 3-8 inclusive in the schools surveyed, as a means of selecting those cases which needed further individual examinations. A considerable number were then examined individually with the Stanford Revision. The schools surveyed were the rural and village schools in nine out of the twelve townships in the county, and the schools of Bloomington. The average percent of feeble-mindedness for the schools outside of the county seat was found to be 6.4, averages for individual townships varying from 2 percent to 20 percent. In the schools of Bloomington 2 percent were feeble-minded.

Discussion of Surveys

It will be seen that these surveys show a wide variation in the proportion of feeble-mindedness in the population.[11] The smallest percent appears in a study which was not primarily an effort to discover all cases of feeble-mindedness, but to learn under what conditions the known cases live, and the largest percent, in a study where children who were likely to be cases of mental deficiency were selected by the use of group tests and teachers' estimates and were given a careful individual examination with standardized mental tests. In most instances, the smallest ratios found are in surveys in which information in regard to the numbers of cases was secured from individuals and agencies not especially qualified to diagnose mental deficiency. The more careful and intelligent these observers were, the more certain they would be to hesitate to classify as feeble-minded, any cases which were at all doubtful, thus missing the greater number of higher-grade cases which to the untrained observer so frequently appear normal.

It is an oft-repeated experience that well-intentioned individuals prefer to excuse indications of mental inferiority in others on the basis of ill-health, poor home care, language difficulty, lack of opportunity, etc., unless the case is a troublesome or delinquent one. This is especially true in the case of teachers' judgments even in cases where a child is several

[11] See Miner, *op. cit.*, for reviews of other surveys. Also Mental Hygiene, Vol. I, No. 1, on a report of the survey of a section of Baltimore by Dr. C. Macfie Campbell.

years retarded pedagogically. The age-grade status of the child, the most important criterion at the teacher's disposal, is the most frequently disregarded. Since, as is well known, these higher grade feeble-minded constitute the majority of the total number of feeble-minded, their omission would materially reduce the percent of feeble-minded. The writer does not doubt, that the various areas surveyed differ somewhat in their respective percentages of mental deficiency, but it is also her belief that most of these differences can be explained by the difference in the methods of conducting the surveys and in the standards used by the various investigators in diagnosing mental deficiency. Further discussion of the surveys should serve to bring this out.

In the Ohio survey, in the case of school children, all children who were three years or more retarded pedagogically, together with others named by the teachers, were chosen for individual consideration and the effort was made to predict from the known facts, whether the child would be able to maintain an independent existence when mature. Considering the difficulty of making such a prediction with no fixed standards of comparison, it is surprising that as large a percentage as 1.8 was classified as definitely feeble-minded. But, on the other hand, many children, especially younger children, who are less than three years retarded pedagogically, prove to be definitely feeble-minded. Mental retardation is present in a defective child when he enters school whether he is entered by his parents at the normal age or not. A defective child would have to remain in the first grade three years before he would be considered a suspect if the above basis was used for selection, whereas, at the age of six he was practically as mentally retarded as at the age of nine. It is true that some children were finally classified as feeble-minded where the pedagogical retardation was less than three years, but the proportion is very small and from the description, the cases were evidently very low-grade, at least to be attending public school. So that it seems safe to conclude that many higher grade cases of defect must have been overlooked with the method of selection used.

It is often stated that there is a marked tendency on the part of the lay public not to report a case as feeble-minded unless it is quite low-grade or unless it is anti-social in its reactions. In the New Castle County survey by the Children's Bureau, 1100 cases were reported to the investigators as possibly feeble-minded. Only about a fifth of this number was classified as positive cases and an additional one-third

as questionable cases. It is stated that a number of cases was dropped because of the difficulty of diagnosis or because of lack of sufficient evidence to classify them as defective. In the report of the school cases which were defective, the least retarded cases in the various grades have been noted above. It is evident from an examination of these, that many cases of milder defect have undoubtedly not been included. Miss Lundberg clearly recognized that the numbers quoted did not include all the mentally defective cases and that they could only represent the minimum of all cases in the county.

The New Castle County survey by the United States Public Health Service confined itself entirely to the school children of the county and, as has been noted, gave a considerably larger percent of mentally defective and probably mentally defective children. In a number of the surveys reviewed, it is evident that there must have been a good deal of variation in the judgments of different investigators engaged in the same survey. Although the fact is not stated, it is evident that a number of investigators must have been engaged simultaneously in this survey as 3793 children were studied in four months. A part of the preliminary examination given consisted of "questions suited in a general way to the child's age and school grade," and a part of the secondary examination consisted of tests and exercises that could not help but vary a great deal in the case of different examiners when these tests and questions had not been previously standardized or were incapable of standardization. Even though the percentage quoted here is higher than some of the others, it is evident that it is too low when cases which have a chronological age of 15 and a mental age of 10.6, or a chronological age of 14 and a mental age of 9.8 are accepted as normal because so considered, in spite of the above evidence, and even though the fact that the Goddard Scale measures somewhat short for these ages is taken into consideration.

A consideration of the basis used for classifying children as exceptionally retarded in the survey of Porter County, Indiana, will suffice to show the possibility of overlooking cases of defect. Although perhaps not exactly comparable for the Goddard Revision, it will bring out the extent to which the levels at which a child was considered exceptionally retarded varied for the different ages, if the mental and chronological ages quoted are changed into intelligence quotients for each age. A six year child with 67 percent intelligence was so classified, a seven year child with 71 percent, an eight year old with 63-75 percent, a nine year old with 67-78 percent, a

ten year old with 60-70 percent, an eleven year old with 64-73 percent, a twelve year old with 67-75 percent, a thirteen year old with 61-69 percent, a fourteen year old with 64-71 percent, a fifteen year old with 60-67 percent, a sixteen year old with 67 percent, and a seventeen to a twenty year old with 73 percent, when fifteen years is used as the divisor for the ages above fifteen. Further standardization of the Binet-Simon tests and of the earlier revisions has shown that the tests for the lower ages are too easy so that consequently the ratings of the younger children by the Goddard Revision are too high. Without considering this fact, it is evident from the above percents, that it is highly probable that numbers of cases of feeble-mindedness were not so classified, and that there was much possibility for variation in judgments especially for those ages, 8-15, for which two mental ages are given.

The Maine survey did not attempt an enumeration of all cases, and only studied a very small proportion individually, which is sufficient to account for the small percentage of feeble-mindedness which the investigators found.

There are several outstanding features of the survey of "X" County, California. The first of these is the careful selection of cases by the use of three criteria, (1) teachers' judgments of school work, (2) teachers' estimates of ability, and (3) the amount of pedagogical retardation; second, it would seem that a large enough proportion was examined individually to include all suspected cases; third, these cases were examined by a carefully standardized scale and by investigators whose preparation was directly in this field, and who had had sufficient training in this field to make their results authoritative, especially when verified by Dr. Terman. Some of the other surveys have had one or more of these features, but the writer believes none of them has fulfilled so well the requirements of a well-conducted survey. Dr. Terman considers the percent of feeble-mindedness found in this study "astonishingly high," which it is when compared with the results of others, and concludes that "the ratio in 'X' County indicates an exceptional condition." The writer would not venture to deny this conclusion, but would again suggest that the disparity in the results of this and other surveys is due, in large part, to the differences in the methods of conducting them.

The Presseys in the survey of "X" County, Indiana, went one step further in improving upon the method of making a mental survey, by introducing the use of group tests. By their use an approximate mental rating was obtained of every

child in grades 3-8 inclusive in all the schools surveyed, thus making the chances even less of not detecting all cases of feeble-mindedness. Group tests were not given, however, in grades 1 and 2. It is in these lower grades that the probability of error in the teachers' selections is greatest. The tendency of the schools is to keep the younger children under observation for several years, before reporting them as mentally inferior, and then to report a child who is noticeably over-age only in case he is also doing inferior work. In many instances, feeble-minded children in the first and second grades are not reported to the examiner because they have not been in school long enough to become sufficiently retarded to be regarded as likely cases of mental deficiency. In other cases, where they are sufficiently over-age for their grade so that they are able to do passing work, this latter consideration is sufficient to exclude them from a list of supposed mental defectives. For the greatest reliability in results, it is necessary to have a means of checking the teachers' selections from these grades as well as from the upper grades.

Methods and Scope of Minnesota Surveys

Simultaneously with the beginnings of the survey of "X" County, Indiana, the Research Department of the Minnesota School for Feeble-Minded began a survey of "X" County, Minnesota. One of the first steps in the preparations for making this survey was the devising of a group test scale by Miss Frances Lowell who began the work of the survey.[12] This scale was revised slightly by Miss Maude Merrill and during the past three years has been further revised and extended by the writer, all of the work on the scale being done under the general direction of Dr. F. Kuhlmann. At present the scale contains groups of tests applicable to each age from five years to maturity. As the survey is not yet completed, nor the group tests ready for publication, it is not intended here to give the results of the county survey as a whole nor to present the group test scale, but to present the methods and results of the surveys of: (1) a group of 23 city school systems in all parts of Minnesota having grade populations varying from 162 to 2640; (2) a group of three city school systems, "A," "B," and "C," having grade enrollments of 962, 1301, and 1490, respectively; and (3) a group of 35 rural and small village schools with a total enrollment of 757. In the first group, selections of cases for individual

[12] Lowell, Frances, A Group Intelligence Scale for Primary Grades, *Journal of Applied Psychology*, Sept., 1919, Vol. III, pp. 215-247.

examinations have been made by the teachers in their respective schools; in the last two groups, the method of selection for individual examinations has been by the use of the group tests. In the discussion, these two methods of selection will be referred to as the 'group-test-selection' and 'teacher-selection' methods. The study comprises a total of 1490 individual examinations and 3524 group test examinations.

Minnesota provides state-aid for special classes for subnormal children in the public schools throughout the state. These classes are limited to an enrollment of fifteen and to children having an intelligence quotient of 50-85. Examinations of candidates for these classes are made by the Research Department of which the writer is a member. The method of making a survey of a school system to determine those eligible to the special class consists first of selection by the teachers of all cases in their own rooms which they consider sufficiently retarded to be eligible to the special class. Preliminary information blanks, supplied by the state supervisor of special classes, which call for the complete school record of the child, information in regard to the parents, home care the child receives, etc., are then prepared. In addition to this, each child is given a medical examination. These data are prepared before the mental examiner arrives and are made available before the mental examination takes place. The plan followed is to test all cases selected by the teachers whether the class or classes can accommodate them or not. In the opinion of each school system, the survey has included all cases suspected of mental deficiency and is complete. The examinations in the group of 23 schools mentioned above have been made in connection with such surveys for special classes. They represent a selected group of schools which have been surveyed in this manner. In recent years group tests have been given in a number of places as an aid in selecting the cases for individual examination; in others, the special class teacher, who has often taken a short course in testing, has examined the children in order to have a more selected list for the state examiner. Since such cases would not of course be a fair sample of the 'teacher-selection' method, only those schools where no preliminary mental testing has been done, and where examinations were made for the first time have been included in the group under consideration.

City "A" and all the schools in group 3 are in "X" County and have been surveyed as a part of the county survey. City "A" is the largest city and the county-seat of the county. The group of 35 rural schools is not a selected group except in

that these schools are the most recently surveyed of the rural schools of the county. Cities "B" and "C" of group 2 are representative cities in other parts of the state which were surveyed, (1) as a means of comparison with city "A," and (2) as a means of comparison of the 'group-test-selection' method with the 'teacher-selection' method.

All examinations, both group and individual, have been made by the members of this research department. The Kuhlmann Revision of the Binet-Simon tests has been used for individual examinations.[13] As reliable results can be obtained through the use of this revision as are at the present time possible, because of the thoroughness and accuracy of its standardization, the large number of tests for each year from 3-15, and its continued improvement through its several revisions. An advantage has been that an abbreviated scale consisting of largely 'non-verbal' tests has been derived from it for use on cases who suffer from a language handicap or who, because of excessive timidity, would not respond as readily to the tests which require longer verbal responses or the formulation of answers. Comparison of this 'half-scale' examination with the full-scale examination of a large number of cases of the above description showed that the score obtained on the 'non-verbal' tests alone was always higher than that on the entire scale. In a considerable number of cases in the rural schools, this 'non-verbal' examination was used in place of the complete examination. It should be borne in mind in the consideration of the rural school results that the possibility of error in these test results is in the direction of being too high rather than too low, which would have been the reverse had the full-scale examination been given to these cases.

As the medical and social data necessary to the final diagnosis of borderline cases have not been completed in the case of the rural school examinations and those of cities "A," "B," and "C," the cases examined have been classified on the basis of the individual mental examination results alone. Cases with an I. Q. of 50-74 are classified as morons; 75-84, as borderline; 85-94, as dull; 95-104, as average normal; 105-114, as bright, etc. This classification is of course somewhat arbitrary and it is possible that some cases are included in the moron group, who in the light of the completed data would not necessarily be classed as feeble-minded in the narrower sense,[14] but on the other hand, it is certain that

[13] Kuhlmann, F., Handbook of Mental Tests, Warwick & York, Baltimore, 1921.

[14] That is, requiring institutional care or supervision because of failure, due to lack of intelligence, to pass the social test.

many cases testing 75 or a few points above would, with extended information, be classed with the definitely feeble-minded. It is possible that these two changes might compensate each other. A recent extensive study by Dr. Kuhlmann has shown that there is a general tendency for the intelligence quotients of defective children to decrease as they grow older.[15] Those cases who in childhood have an I. Q. of 75 or a few points above or below will in all probability not maintain this I. Q., and by the time they are adults, it will be somewhat lower. Dr. Kuhlmann has concluded that the chances are practically nil " that such adults (below an I. Q. of 75 or 70) will be able permanently to make an independent, honest living, without supervision and guardianship, under any and all circumstances they are likely to meet in their lives."[16] Other considerations aside, all those with an I. Q. of 50-74 are as mentally inferior as the high grade feeble-minded. They constitute the group which is the greatest problem to the public school and society, and to which school and society are problems of like magnitude. It is of importance to know how numerous this class is, in order to have a more comprehensive program to deal with the difficulties incident to its members. To attempt to minimize the importance of the problem, by presenting data which are only a partial picture of the conditions, can only have results disastrous to the efforts of those attempting to deal with conditions as they actually are.

Comparison of Two Methods of Selection in City Schools

TABLE I

	Total No. Cases	No. Indiv. Tests	Percent. Indiv. Tests	F. M.	Border Line	Dull	Av. or Above	Percent of Total F. M.
'Teacher—selection'	20,877	911	4.4%	50%	25%	18%	7%	2.2
'Group-test—selection'	3,001	121	4.4%	67%	25%	8%	0%	3.1

Of the total enrollment in the 23 cities 911 cases or 4.4 percent were selected by the teachers for individual examinations. In the three cities surveyed by the group test method, the proportion of individual examinations was considerably greater. The total number examined in the latter was 315 or 10.5 percent of the number given the group test. In order

[15] Kuhlmann, F., The Results of Repeated Re-Examinations of 639 Feeble-Minded Over a Period of Ten Years, *Journal of Applied Psychology*, Sept., 1921, Vol. V, pp. 195-224.

[16] Kuhlmann, F., Determination of Feeble-Mindedness as Related to the Courts, Minnesota State Board of Control, Saint Paul, 1920.

to have the results of the two types of surveys more nearly comparable, the individual ratings of the 4.4 percent who received the lowest ratings in the group tests are compared (Table I) with the 4.4 percent chosen by the teachers in the 23 cities. In the latter group, 75 percent rated feeble-minded or borderline in the individual test, and the remaining 25 percent rated dull, average, or slightly above the average. The cases rating feeble-minded constituted 2.2 percent of the total school enrollment. Out of the 4.4 percent lowest in the group tests, 92 percent rated feeble-minded or borderline in the individual test and 8 percent dull, with none rating above this. This lowest 4.4 percent in the group tests yields a percentage of 3.1 percent of the total rating feeble-minded. Out of a hundred children, selected by each of the two methods, for individual examinations the group-test-selection method discovers 17 more cases of mental deficiency, and eliminates, on the other hand, 17 individual examinations of cases too high to classify as either feeble-minded or borderline.

The results of the completed survey show that the percentage considered above is not large enough to reveal all cases below an I. Q. of 75. As has been stated, 10.5 percent of the total number reached in the group test were also given individual tests. It is obvious that the larger the percentage tested, the greater becomes the possibility of error in selecting the cases, because the scores approach more nearly the average and a smaller gain in the individual test will put the case into the dull or average group. In spite of this fact, the larger percentage of 10.5 percent compares favorably with the 4.4 percent selected by the teachers. Out of the total number of 315 cases examined individually, 77 percent classified as feeble-minded or borderline, and 23 percent as dull or average, compared with 75 percent and 25 percent, respectively, for the other method. The percent of cases rating below an I. Q. of 75 were: City "A"—4.5 percent; City "B"—3.5 percent; and City "C"—4 percent, an average of 4 percent for the three taken together. The percents quoted for cities "A" and "C" are low because the former system already had one special class and the latter, two special classes organized before the group test survey. The cases in these classes are not included in these results. The Presseys found that all the cases of mental deficiency were included in the lowest 10 percent in the group test. This agrees very closely with the percent it was found necessary to examine individually in the above schools. With further improvement in the group tests it may be possible to reduce this percentage somewhat.

The question arises as to how representative these results are likely to be of the population as a whole. Any study, to be representative, must include a large group of unselected cases. The only place where it is at all possible to find such a group is in the public schools, especially in the lower grades. Even in these selection has begun to occur because low grade cases have not been entered at all or have been excluded shortly after their entrance. The percent of these is small, however, and their omission does not affect the results appreciably. In the upper grades, elimination takes place, the lower cases here dropping out after they have passed the age for compulsory school attendance. But, on the whole, the sampling secured here is the most representative that is possible, and the distribution of grades of intelligence is undoubtedly approximately the same as would be found in any equally representative group at any age. As Miner has indicated,[17] the fact that mental deficiency is not a condition overcome with age, but a permanent state, assures closer agreement between the percent of feeble-minded in the school population and the general population than most investigators have been accustomed to quote, even when allowance is made for the greater death rate among the feeble-minded.

At the time of its use in these surveys, the group test scale was by no means as accurate as it will be when finally standardized. When the scale is in its final state, it is expected that the efficiency of the group-test-selection method over the teacher-selection method will be even greater. Because of the individual variations which are always present in the reactions of certain types and ages of children in a group test and an individual test, there will always be a certain amount of variation between the results secured by the two types of scales, however carefully standardized, which it will be impossible to control. For this reason the group test can never hope to supplant the individual test nor give a final judgment on a case.

This added efficiency in the selection of cases is only one of the advantages of this type of survey. The great saving in time and expense must not be overlooked. The more accurate and efficient selection by the group-test-selection method has been possible after only a fraction of a day spent in each schoolroom, whereas, in the case of the teachers' selection, it has usually been necessary for a child to repeat one or more grades before the evidence is considered sufficient to place him in the suspected group. Individual mental test results

[17] *Op. cit.*

of a number of children ranking highest in the group tests indicate that the method is just as efficient for the selection of the most superior children for further study.[18] Comparisons of the averages for different buildings and grades are possible and often reveal unsuspected and startling discrepancies between the average levels of ability of the same grade in different buildings or of different grades in the same building. The problem of the teacher who has, on the average, less well-endowed pupils is better understood, and the failure of these pupils to measure up to others in pedagogical tests need not then reflect discredit on the conscientious teacher, as is otherwise apt to be the case. Cases have been reported of certain teachers receiving additional recognition because of the rank of their pupils in pedagogical tests. Intelligence tests of these same pupils have shown that, in reality, their attainment was below what it should have been when their native ability was taken into consideration.

Differences in Grade and Building Averages

It was expected that the average levels of ability in the various grades and buildings would differ somewhat. In certain instances, however, the contrasts between grades were so pronounced that a study was made of the results for the three cities "A," "B," and "C" to determine how general and how extended such differences were. The results of this study are so suggestive that it was thought advisable to include them in this report. The variation in average ability of the different grades and buildings and the variation in the distribution of the brightest and dullest children of each system among the different buildings will be shown.

With the group test scale that has been used, it has been possible to secure a mental age and an I. Q. for each child taking the group test. A comparison of the results of the individual test and the group test for the same children has shown that the group test measures slightly short, that is, the mental age secured by the group test is, in most instances, lower than that secured by the individual test. In order to have the average group test rating more nearly correct, the group test scores have been corrected, so that the average score for each chronological age is 100, before the averages for the grades and buildings have been secured. This assures the accuracy on the average of the group test results. As a check on the accuracy, the average 'corrected' group test

[18] See Book, *op. cit.*, for data on this point.

score was compared with the average individual test score for all the children in one of the cities ("B") who were given individual tests. The former score was found to be .3 of one point higher than the latter, so that the averages can be accepted as reliable estimates of the average abilities of the various grades and buildings.

TABLE II

AVERAGE SCORES FOR GRADES AND BUILDINGS

Schools	Grade I	Grade II	Grade III	Grade IV	Grade V	Grade VI	Bldg.	City
City "A"—I	107	103	104	101	101	105	104	
II	98	107	100	103	105	107	103	
III	90	97	99	107	107	104	101	
IV	95	93	96	95	89	105	96	
V	85	99	93	98	94	94	94	
								100
City "B"—I	96	106	102	100	103	104	102	
II	95	98	101	99	101	104	100	
III	93	97	98	96	94	100	96	
IV	97	95	94	97	94		95	
								98
City "C"—I	98	104	102	101	107	97		
			96				101	
II	104	102	96	97	97	101		
	92			93	96	92		
					92		97	
III	88	99	101					
		94					96	
IV	96	97	95	94	96	94	95	
V	93	101	97	102	96	93		
				88	93	87	94	
								97

TABLE III

AMOUNT OF VARIATION IN SAME GRADE IN DIFFERENT BUILDINGS

	Grade I	Grade II	Grade III	Grade IV	Grade V	Grade VI
City "A"	22	14	11	12	16	13
" "B"	4	11	8	4	9	4
" "C"	16	10	7	14	15	14

TABLE IV

AMOUNT OF VARIATION IN DIFFERENT GRADES IN THE SAME BUILDING

	School I	School II	School III	School IV	School V
City "A"	6	9	17	16	14
" "B"	10	9	7	3	
" "C"	11	12	13	3	15

In Table II are given the average scores for each grade, building, and city. From Table II, Tables III and IV have been derived, summarizing the range in scores in the different grades and buildings. In City "A" there is a variation of 22 points in the average scores of the first grades in the different buildings. The significance of this difference is more clearly seen if the difference in I. Q. is translated into the corresponding difference in mental age. If the assumption is made that the average chronological age in the first grade was seven years,[19] the average mental age for Grade I in School V was 5.95, and for Grade I in School I was 7.49, an average difference of 1.54 years in mental age between two grades in the same system from which the same results are expected. Assuming the average chronological age for the second grade to be eight years, for the third, nine years, etc., the same comparisons can be made for the other grades. Perhaps the large differences between the same grade in different buildings are not so unexpected. It is more surprising to find that almost as large and as frequent discrepancies occur in the different grades in the same building. Tables (II-IV) should serve to bring out forcibly the fact that it is not sufficient to have a general notion that one grade is better or poorer than another, but that it is necessary, for the best results, to have an accurate measure of how much above or below average a grade or building is.

Distribution of Extremes in Ability Among Grades and Buildings

As a further means of comparison of the different systems, the highest 10 percent and the lowest 10 percent in ability were selected in each system, and the distribution of these best and poorest among the different buildings was studied. These results are shown in Plates I and II. In Plate I, the percent of each building's enrollment which is in the 10 percent lowest and 10 percent highest of the cities is shown. If these were distributed uniformly each building would have 10 percent of its enrollment in each of these groups. However, the distribution is far from uniform. In city "A," School IV has four times as large a percent of its enrollment among the 10 percent lowest than School II, and on the other hand, School I has over three times as large a percent of its enrollment in the 10 percent highest than Schools IV and V have. The contrasts are almost as striking in cities "B" and "C."

[19] All of the surveys were made at or after the middle of the school year.

PLATE I

PER CENT OF EACH BUILDING'S ENROLLMENT IN THE 10% LOWEST AND HIGHEST OF THE CITY

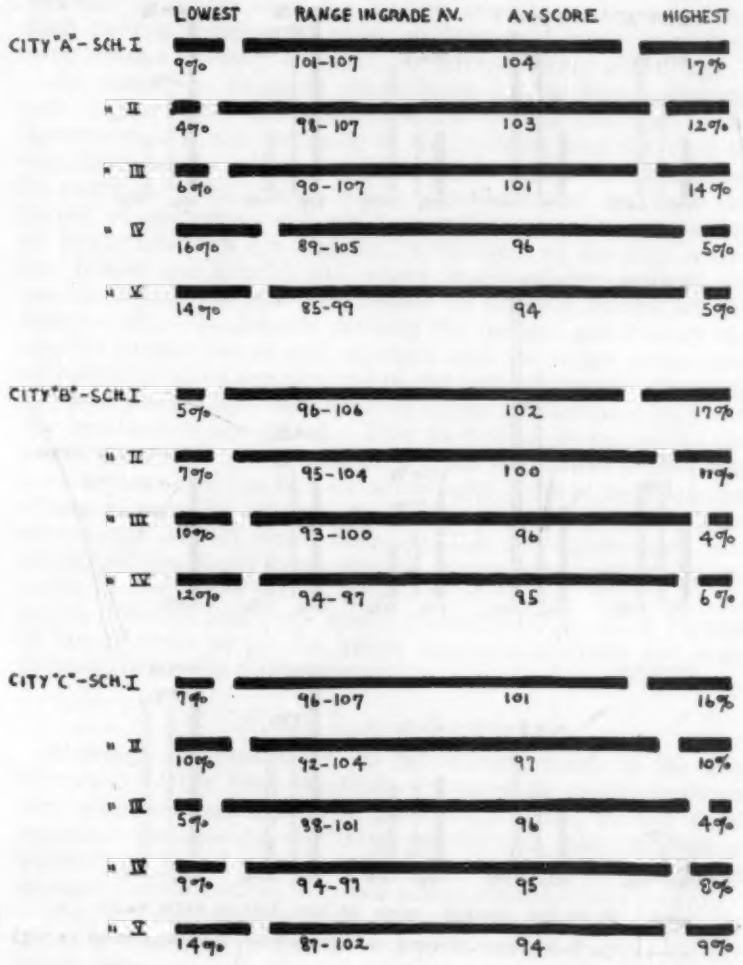


PLATE II

DISTRIBUTION OF LOWEST AND HIGHEST 10% OF EACH CITY AMONG THE DIFFERENT BUILDINGS

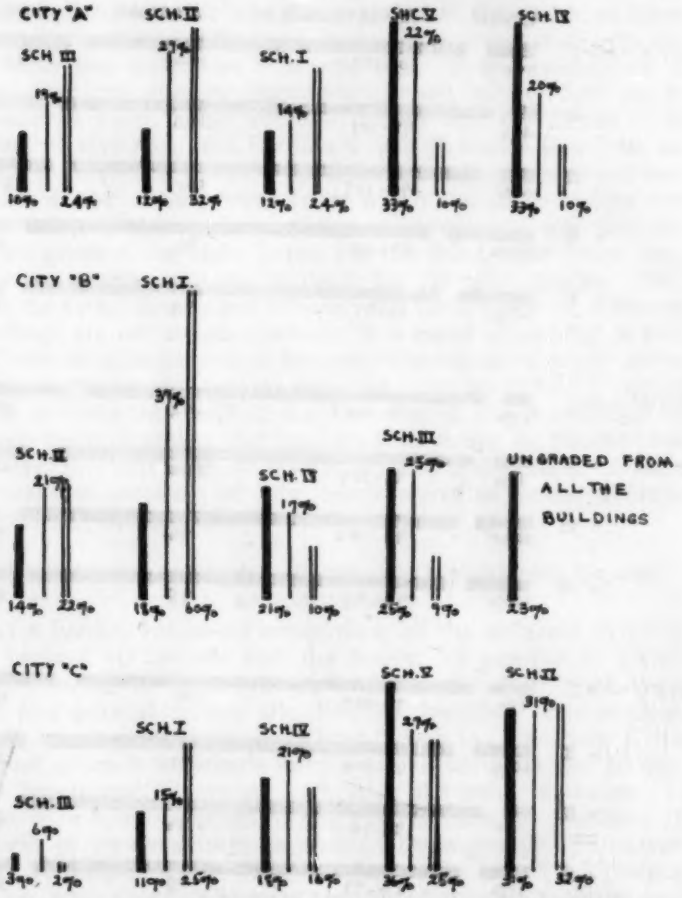


Plate II gives a comparison from another point of view. The distribution among the buildings of the children ranking in the best and poorest groups is here compared with the percent the building would have if the best and poorest were distributed in proportion to the size of the school. In general the schools which have the highest average scores have less than their proportion of weak students and more than their proportion of strong students, the reverse being also true.

As would be expected the schools in the better districts rate higher than the schools in the poorer districts. The agreement between the rank of the building and the type of neighborhood in which the building is found is very close. In every case, the schools which rate the lowest and have the larger proportion of dull children and the smaller proportion of bright children are situated in the part of the city where the homes are poorer and where many of the fathers are unskilled laborers and are unable to provide homes of the better sort. The schools ranking the highest and having the smaller proportion of dull children and the larger proportion of bright children are situated in the part of the city in which the professional and business men of the city live and in which the best homes are found. This fact often gives rise to the erroneous belief that the children from the poorer homes rate lower because of the lack of advantages in their environment, when, as a matter of fact, the poor environment and lack of advantages are in most instances the accompaniments and results of the lower mentality of the parents instead of the cause of the lower mentality of the children. The contrasts in the different buildings would undoubtedly be much greater in larger cities where the better residence districts and slum districts are more differentiated.

Rural School Surveys

Although the percentage of feeble-mindedness in the city schools is higher than would be expected by many, it seems low when compared to the percentage found in the rural schools. Because of the large number of cases of feeble-mindedness found in the latter, the writer must confess some hesitancy in presenting the rural school results.

The work in these schools was done in the fall when the weather was still sufficiently warm to permit the use of the small entrance hall to the school-house for the individual examinations, thus avoiding disturbances, interruptions, and lack of response on the part of the subject because of the presence of others. The examiner arrived at about the time

school started in the morning and the group test was given immediately, everyone taking it at once. The tests for the youngest age groups were given first and the explanation was made that some easier things would be given first so that the younger children could do them too. As the younger children automatically dropped out because of the increasing difficulty of the tests, their books were taken up and the teacher gave them seat work to do. As this procedure lengthened the process of giving the group test considerably, a recess was given when the upper age tests were reached. During this time the examiner looked hastily through the books and took up those in which the subject had obviously gone beyond his limit. The rest of the tests were then given and the remainder of the morning was utilized for scoring the results. The children obviously enjoyed the tests and were stimulated to their best effort when they were told that the same ones were being given in all the schools of the county and that one of the objects was to see which school could do them the best. At noon the examiner ate her lunch with the teacher and children and often joined in their games, so that by the time school was called in the afternoon any feeling of timidity had worn off. The cases chosen for individual examinations were plainly envied because of this distinction. The explanation was made that only a limited number could be taken because of the time factor. Where ever there was an indication of language difficulty (some of the schools were in Bohemian districts) or of hesitancy in responding to the verbal type of tests, the 'non-verbal' examination already referred to was given. It would seem that conditions were as favorable as those usually obtainable. In smaller schools where only one or two individual examinations were necessary, it was possible to finish the school in a day, in larger ones, it was necessary to return for a second day or fraction thereof. Only one school had more than one teacher. This one had three teachers and was given three days time.

In the 35 schools, individual examinations were given to 35 percent, 183 cases, of the total number taking the group tests. Five of these schools had no cases with an I. Q. below 75, nine had from 1-10 percent, 13 had 10-25 percent, and eight had over 25 percent of those present rating below an I. Q. of 75, an average of 15 percent for all the schools taken together. Besides this an additional 11 percent rated between 75 and 84 in the individual examination. In two schools no individual examinations were necessary, in fifteen schools more than 35 percent of those present were examined, nine of these

having over 50 percent of those present examined. Eleven of the schools had 66 percent of all the cases below an I. Q. of 75.[20] This large percentage of individual examinations was necessary because of two things; the large number of cases testing below 75 and a few points above, and the fact that the rural school children are more accustomed to being dealt with individually or in very small groups, and therefore their performance in the larger group was often a good deal poorer than it was in the individual test. This was especially true of the younger children.

The large percentage of cases below an I. Q. of 75 may signify one of several things. It may show, as Pressey has asserted, that the standardized intelligence scales in use have features that are distinctly urban in nature and that rural school children should not be judged by them on the same basis as city children. Another conclusion possible is that these schools are not representative of rural schools as a whole, but that their extremely high percentage of cases of low mentality represents an abnormal condition and is disproportionately high for rural communities in general. On the other hand, it may serve as conclusive proof of the generally-held opinion that a higher grade of mentality is required to maintain an existence in an urban community than is required in a rural community.

In regard to the first, an examination of the tests in the scale used reveals only a very few tests which might be passed with greater ease by a city child than by a rural child of the same mentality. But even though this were true, it is possible that tests which might serve as a truer index of a rural child's mental capacity could be substituted for some of these, and that the lower rating of the rural children may be due, in part, to

[20] It has been noted that the average percent of feeble-mindedness for the townships in the survey of X County, Indiana, varied from two to twenty. In these 35 rural schools in Minnesota, the variation was from 0 to 50 percent for schools, and from 6 to 20 percent for townships. These large discrepancies in the results within the same survey might seem to weaken the writer's contention that the differences in the findings of the surveys discussed were due largely to the differences in the methods used in conducting the surveys and in diagnosing mental deficiency. The writer has not attempted to show that these differences were always or entirely due to these differences in methods. Local differences are granted and expected. These large differences in neighboring schools and townships emphasize the necessity of obtaining a sampling from a large enough area to represent an average of conditions. All of the surveys discussed have fulfilled this condition. It is where averages from such representative samplings vary from .212 percent, .324 percent, and .44 percent to several times this proportion that the question of method is pertinent.

the omission of certain types of tests rather than to the unfairness of tests already included. At first thought it would seem that this could only account for a small percentage of the difference between the city school and rural school results. However, closer inspection reveals that this factor might influence the percent rating mentally deficient very materially. Of the total number rating below an I. Q. of 75, 38 percent fall in the range 70-74; 53 percent in the range 60-69; 7.5 percent in the range 50-59; and 2.5 percent below 50. Suppose the assumption is granted that the tests used measure slightly low for rural school children, and that the limit for arbitrary classification of cases as feeble-minded is dropped five points. Those cases falling below an I. Q. of 70 would then be classified as feeble-minded. This drop of five points changes the percent of feeble-minded from 15 to 9.5, a very considerable reduction. The need of further research along this line is suggested. It is not improbable that further study might prove that the above assumption is justifiable.

The county being surveyed contains fourteen townships. Seven of these are represented in the group of schools under consideration, and only two of these by less than five schools. The results of the completed survey may show that conditions in the other townships are better than in those represented in this study. It is certain that one of the townships, which contains six of the schools out of the 35, is not inhabited by enterprising farmers because of the poor quality and unfavorable nature of much of the land in it. The percents of cases below an I. Q. of 75 in these schools are: 7, 10, 22, 36, 40, and 50. This last school had an additional 30 percent which graded as borderline cases. The enrollment of the school was made up of children from five families. The children of two of these, which comprised all but four of the total enrollment, were cousins. One of the remaining four was a middle-grade imbecile and was almost totally blind, one was a borderline case, and the remaining two were a brother and a sister who were normal. Even if the majority of the cases of mental deficiency in the county were in these 35 schools, the percentage for the county would still be extremely high compared to that quoted by other investigators. It is probable that the county as a whole is not representative of Minnesota.

That this is undoubtedly not a representative Minnesota county is indicated by the 1920 United States census report for Minnesota.[21] This report includes figures on the percent

[21] United States Bureau of the Census, Fourteenth Census of the United States: 1920, Population: Minnesota, Composition and Characteristics of the Population.

of illiteracy in the counties and cities of Minnesota. For the state taken as a whole the percent of illiteracy is 1.8. Sixty out of the 86 counties have less than two percent, 21 have two to five percent, and five have six to eight percent. The highest percent for any county is 7.6 percent, the next lowest is 7.5 percent. The latter percent is that for the county under consideration. Of the eleven Minnesota cities having a population over 10,000, six have less than two percent of illiteracy, and three have from two to five percent. The percents for the remaining two are 7.7 and 16.1. The latter of these is for City "A" in this report.

These two considerations, that the mental test ratings do not do justice to the ability of the rural child, and that these particular schools have an unusually large percentage of cases of low mentality, are of importance. The third also bears considerable weight. Results from rural schools would be expected to be somewhat lower than those for city schools. That the difference is so great is indeed unexpected. As has been mentioned, 9.5 percent of the rural school cases fall below an I. Q. of 70. The percent of cases falling below 65 is 4.5; below 60, 1.5; below 50, .4. These percentages seem significant in the light of other surveys. It has been maintained that investigators who have depended upon social agencies, public officials, etc., for their information in regard to cases have discovered only the lower grade cases. That the percent of cases below an I. Q. of 50, i.e., imbecile or idiot-grade, found in these rural schools so nearly agrees with the results of surveys conducted in the above-mentioned manner seems to substantiate this belief.

In one or two surveys where tests of various kinds have been used as a partial means of diagnosis, it is evident, that in case the test result has not agreed with the investigator's personal judgment of the case, the latter has had the greater weight. There can be no question, that in such cases, a large number of the higher-grade feeble-minded, who are so difficult to classify, even for an expert in this particular field, have been dropped as not feeble-minded. It has been stated that this would be true in the absence of any mental rating. It seems a safe guess to say that the majority of cases, especially children, above an I. Q. of 60 or 65 would not be classed as definitely feeble-minded unless characterized by physical stigmata or unfavorable temperamental traits, by anyone not especially trained and experienced in the field of mental diagnosis, particularly when these cases are seen in surroundings where standards of achievement are not so high. It is alto-

gether possible that these same cases, when found in a rural community, and when judged by the social criterion, would not necessarily be classified as feeble-minded, as this term is legally applied in Minnesota.[22] It is certain that a much smaller percent of these cases will need institutional care or state guardianship than of the cases of the same mentality in city schools. The 1.5 percent rating below an I. Q. of 60 in the rural schools seems to represent the group which would be classified as unquestionably feeble-minded had these schools been surveyed in the manner which has been indicated above. The 4.5 percent rating below 65 may be considered to comprise those cases most likely to become socially mal-adjusted and incompetent to maintain a livelihood independently. The 10.7 percent rating just above these are very likely the group which would not ordinarily be classed as feeble-minded because of their apparent normality when children, and because of the fact that in most instances, the simplicity of their environment and the relative ease of earning some sort of a living do not make greater demands on the adults of this group than their limited intelligence can meet.

The social investigations have been made for about one-fourth of the cases rating below an I. Q. of 75. The personal and family history obtained and the prevailing home conditions have confirmed the mental ratings in practically every case. In no case was there entire absence of significant features in the additional information thus received. It seems significant that in most of the schools surveyed, the teachers, a number of whom had taught in other parts of the county and state, did not recognize that they were dealing with an abnormal situation, or that they had cases of mental deficiency in their schools, although questioning usually brought out the fact that these children were considered 'peculiar,' or 'very slow,' or 'hard to understand.'

Three Surveys Compared

The results for the rural schools of "X" County and those for City "A" of this county have been given separately. The combined results for these schools yield an average of 8.8 percent of cases below an I. Q. of 75. As this is drawing the line higher than was the case in the surveys of the counties in California and Indiana, the percent of cases below an I. Q.

[22] "The term 'feeble-minded persons' in this act means any person, minor or adult, other than an insane person, who is so mentally defective as to be incapable of managing himself and his affairs, and to require supervision, control and care for his own or the public welfare."

of 70 was found. City "A" had 2.6 percent of its cases below an I. Q. of 70, the rural schools 9.5 percent, with an average of 6.1 percent for all the schools combined. These latter percents are comparable with those of Drs. Terman and Book. Dr. Terman classified cases as definitely feeble-minded in case they were below an I. Q. of 70. Dr. Book has not stated where he drew the line in classifying his cases, but since the Stanford Revision was used for individual examinations, it is presumable that it was not drawn above 70.

TABLE V

	City	Rural	Average
California	3.5	4.9	4.24
Indiana	2.	6.4	
Minnesota	2.6	9.5	6.1

The results for these three surveys are given in Table V. Unfortunately the average percent for the rural and city schools is not given for the Indiana Survey. From the two percents given, it can be estimated at about four percent. That the results for these three surveys should agree with each other much more closely than they do with the results of any of the other surveys reviewed above would be expected, as there are several points of similarity in the methods employed. The chief of these points of similarity are: (1) an accurate objective standard was used to diagnose mental deficiency; (2) a sufficiently large percent of individual examinations was made to warrant the assumption that all cases of mental deficiency had been found; (3) the surveys were under the direction of specialists in the field of psychology and mental diagnosis; and (4) the examinations were made by individuals trained in psychology and in the giving of mental tests. The chief difference in the methods was in the selection of cases for individual examination, group-tests not being used in the California Survey, nor in grades 1 and 2 in the Indiana Survey. Because of the above-mentioned points of uniformity in procedure, it seems safe to assume that the variation in the percents of feeble-mindedness found in these studies is indicative of the extent to which the localities differ in their respective proportions of mental deficiency.

Summary

1. There is much lack of agreement in the results of the various surveys made to determine the percent of feeble-minded in the general or school population. The actual percentages quoted in the surveys which have been reviewed vary from .16 to 6.4.

2. Most of this lack of agreement is due to the differences in the methods of conducting the surveys and to the variation in the methods and criteria used to diagnose feeble-mindedness.

3. Surveys of school systems in which the teachers have selected the cases suspected of feeble-mindedness give a percent of 2.2 of the total school enrollment which rates as feeble-minded. Teachers' errors are in two directions, selecting children who do not classify as feeble-minded and overlooking others who are feeble-minded. The chief cause of these errors is the failure of the teacher to take into consideration the age-grade status of the child.

4. The most complete type of survey is that in which carefully standardized group tests are used as a means of selecting all cases which are feeble-minded, these cases being further examined with individual mental tests before a final diagnosis is made. Surveys of this type in three cities of Minnesota yield an average of 4 percent of the number given group tests rating below an I. Q. of 75 in an individual mental test. With this method there is much less error in the selection of cases for individual examination than in the case of the teachers' selections.

5. The group test survey also makes possible a comparative study of the different grades and buildings in a system and reveals large discrepancies between the average levels of ability. For instance, the average I. Q. of the poorest grade in one building was found to be two points higher than the average I. Q. of the best grade in another building in the same system. A study of the distribution of the most superior and inferior in ability among the different buildings is also possible and likewise brings out striking contrasts. In one system of four buildings, one building had 60 percent of the most superior children of the entire system.

6. Mental surveys of thirty-five rural schools yield an extremely high percentage of cases rating below an I. Q. of 75. These schools had on the average 15 percent of the number present for the group test who rated below an I. Q. of 75 in an individual mental test. This percent was reduced to 9.5 when the upper limit for arbitrary classification of cases as feeble-minded was placed at an I. Q. of 70 instead of 75.

7. Combined results for the rural schools and City "A" of this county yield an average percent of 8.8 rating below an I. Q. of 75, and an average percent of 6.1 rating below an I. Q. of 70. These findings approximate those of other surveys employing similar methods.

ON FINDING EQUIVALENT SCORES IN TESTS OF INTELLIGENCE

By Edward L. Thorndike, Institute of Educational Research, Teachers College, Columbia University

With the development of group tests and tests for use with higher levels of intelligence, it is becoming more and more necessary to transmute a score obtained with one test into the score that is equivalent to it in some other test, or into a mental age, which is, as now used, itself really an arbitrary score. We have already three tables of the mental ages corresponding to Alpha scores, one set computed by Yerkes and Yoakum ('20, pp. 133 f.); another, computed by Kohs and Proctor (Proctor, '21 pp. 65-68); a third, by DeVoss ('22, p. 24). Alpha and Beta scores have been equated over a part of their ranges. (Psychological Examining in the United States Army. Part III, Chapter II).

The procedure has been to obtain empirically the average score obtained in one test by those of a certain score in the other, and to use the series of such averages, smoothed to eliminate chance irregularities, as the table of equivalents. That is, the regression line is used. This is the natural thing to do and is probably, often, the most practicable thing to do, but it sometimes subjects the results to an important error. Theoretically it is the right thing to do only in special cases; and when it is done the resulting table of equivalents has a special meaning, which is not, I think, the meaning usually attached to such a table by most of those who will use it. It seems desirable therefore to consider the general theory of the determination of equivalent scores.

Case I. Consider first the case where the two tests measure the same thing, as when an inch scale and a centimeter scale are used to measure length, or when two adequate but different tests measure color blindness. If two tests measure the same thing and if each is perfectly reliable, correlating 1.00 with a repetition of itself, the empirical equating of values described above will be satisfactory, of course, *but there will be no regression.* The two tests will correlate perfectly. All the scores in Test II made by those of identical scores in Test I, will be identical.

In practice, of course, there will be a slight unreliability in each, due to the errors that affect all measures, but this will be, say, one part in a thousand, leaving a correlation between the two tests of say .999+ If a random sampling of lengths from 1 inch to 1 yard were measured by two of our most approved methods at the Bureau of Standards, the correlation would, I fancy, be above .99999 and the regression would be negligible in proportion to the magnitudes measured.

Case II. As Case II consider two tests which are known to measure the same thing,[1] but with a considerable unreliability. Such would be single estimates by the unaided eye of each of such a series of lengths first by one physicist in tenths of an inch and second by another in millimeters, or in a less degree the same measurement by two approved methods, but using only a single determination of each value, and letting these values be subject to errors of a careless copyist. Each of the two measures would then correlate imperfectly with itself, and one would correlate imperfectly with the other. The empirical comparison of estimates of length by the unaided eyes of the two physicists would give a substantial regression. 1 inch would "equal" more than 2.54 cm.; 36 inches would "equal" less than the (36) (2.54) cm. which it really equals. In such a case the regression line is obviously not a safe basis for equating the two scores. If the other regression is used 2.54 cm. would be found to equal more than an inch, while 91.44 cm. would equal less than a yard. In such an experiment with lines from 70 mm. to 90 mm. long, the estimated lengths were such as to make 2.5 inches equal to 68.7 mm. (5.2 mm. too long) and 3.8 inches equal to 91.7 mm. (4.8 mm. too short)!

What the regression line measures in Case II is the "attenuation," the combined unreliabilities of the two tests, the errors. If we had only such data to equate inches and centimeters by, we ought not to take the empirical equivalents as they stood but to correct them by some allowance for their "attenuation" or errors.

Suppose, for example, we know that Stanford Mental Age and Army Alpha measure the same thing. We plot the scores of one against the other for some thousands of cases using a single trial with each. If the correlation is not perfect,—

[1] "To measure the same thing" is somewhat ambiguous. The fact that the two measures do not correlate perfectly may be said to prove that they do not measure the same thing, length, but different things, to wit, length plus certain different causes of errors. This may be admitted. What I mean by the phrase here is, I think, clear.

if there is regression from perfect toward zero correlation, we must infer that this is due to our having imperfect measures in Stanford Mental Age or in Army Alpha or in both. A table of values of Alpha scores in terms of Stanford Mental Age got by taking the averages of arrays in Mental Age for each Alpha score—the regression line—smoothed or unsmoothed, is not a table of *equivalents of the scores;* it is simply a table of *prophecies* of how a person who scores a certain amount in one trial with Alpha will probably be rated in Mental Age if he is given one trial with the Stanford test. It is as if we measured a man's stature in inches by having one person guess that it is 180 centimeters and looking up the average guess in inches made by another man when judging a 180 cm. height.

We can obtain equivalents only after first correcting for the attenuation. This can be done by applying Kelley's formula

$$(\sigma_{\infty} = \sigma \text{ obt.} \sqrt{\text{robt. 1. with obt. 2}} \quad [\text{'21, p. 379}])$$

to the Alpha series and the Stanford series. Then, knowing that the correlation really was perfect, we should fit the two corrected surfaces of frequency one on the other and use the correspondence of Alpha and Stanford that makes the best fit. If the attenuations in the two series were equal we could use the values which made the best fit of the two uncorrected surfaces of frequency, one on the other. If the two attenuations are equal, or if we know how the attenuation is divided between the two series, it should be possible to estimate the true equivalence from the two regression lines in this case, (the regression of Alpha on Stanford and of Stanford on Alpha), if we know that some one point in the relation represents the true equivalence and know which point it is. It is often reasonable to assume that the point where the two regression lines cross does approximately represent the true equivalence. It is at any rate the probably best representative of it that the data afford. A convenient way to do this is, if the two attenuations are equal, to plot the data on a scale such that 1 S. D. of A is the same length as 1 S. D. of B and then draw the bisector of the angle between the two regression lines.

I am not competent to derive a formula for the general case but I hope this article may stimulate someone to do so, at least for the case where the correlation is linear. The division of the attenuation between the two variables would be known either through the correlations ($^{r}a_1$ with a_2 and $^{r}b_1$ with b_2), or through the errors of the determinations, (e.g.,

the average σ true—obtained for a and the average σ true—obtained for b).

So the problems stand as follows:

A.

Given correlation plots (linear) of $^{r}a_1b_1$ $^{r}a_1a_2$, and $^{r}b_1b_2$
(Or in place of the two last given the average mean sq. error of the a's and the average mean sq. error of the b's)
Assuming that $^{r}a\ \infty\ b\ \infty = 1.00$
To draw the regression line for $^{r}a\ \infty\ ^{b}\ \infty = 100$
from the regression lines for $^{r}a_1b_1$

B.

The same, but for a correlation plot of any form for $^{r}a_1b_1$.

Case III. Consider now Case III, where we know that the two measures each have perfect reliability,—perfect self correlation. If then, in the empirical pairing of values, we find a lack of perfect correlation, we know that they *do not measure the same thing.* The regression from perfect correlation is then a measure of the unlikeness of the two things measured by the two tests. We ought not then to speak of one as having an equivalent in the other. The case is then logically like estimating a man's weight from nis height. We can make a better estimate of a man's weight from knowing his height than from not knowing anything about him, but 6 feet does not equal *any* number of pounds. It is both logically and in fact like estimating the total wealth of an individual (a) by the taxes he pays and (b) by the amount he can borrow. Both a and b would correlate with the fact to be measured, but imperfectly, since each would measure different aspects of it. The wealth he pays taxes on is not, in every dollar, identical with the wealth he borrows on. It is almost certain that verbal tests like Alpha measure something which is in part not measured by a non-verbal test like Beta. Consequently, if a million persons were each measured by average score in a dozen Beta tests and by average score in a dozen Alpha tests so as to get almost perfect reliability for each, the two scores would still not correlate perfectly. A certain amount of Alpha ability would not then *equal* any amount of Beta ability. We could use one to *prophesy* the other, as in Case II.

There is a difference between the prophecy made in Case II and that made in Case III. We improve the prophecy in Case II toward perfect equivalence by making each test more reliable, for example, by repeating it several times, and we can infer what the true equivalents will be by the changes that occur with each known improvement in reliability. If

disturbances from practice effects, ennui, etc., are not too great, we may bring any two tests that measure the same thing, (for example, ability to spell, or general intelligence, or accuracy of vision) to a real equivalence. We can improve the prophecy in Case III only by altering the tests to make them measure more nearly the same thing. Nothing that we can do will bring the mean square error of estimate of an adult man's weight from his stature down to one pound. We must use more than stature to do that.

Case IV. Consider finally Case IV, where we have two tests which do not correlate perfectly, partly because neither test would correlate perfectly with a second trial of itself and partly because, even if we had hundreds of trials of each, the two averages would not measure quite the same thing. Before speaking of equivalent scores we must alter the tests until they measure the same thing and then correct the empirical equivalents so as to free them from the influence of "attenuation."

At the present time the commonest problems will be to determine the Stanford Mental Ages or the Army Alpha scores equivalent to scores in various group tests. These problems present Case IV with the added complexity that we do not know how much of the reduction from perfect correlation is due to the tests being measures of different traits, and how much to their respective "attenuations." There is, however, no easy way out of the difficulty. True tables of equivalents cannot be computed until we know the unreliability of each test, and the correlation between the two tests when freed from their unreliability (by increased length, repeated measurements, or otherwise). Unless the correlation is then perfect we cannot compute genuine equivalents of units of measure, but only prophecies of scores obtained by persons.

The facts presented here hold, of course, when the two tests to be equated are put in relation by the relation of each to a third variable, such as chronological age. In fact, the error due to taking the empirical averages of arrays (i.e. the regression) is increased in so far as the chronological ages are subject to errors or are grouped coarsely, as by years or half years.

References

De Voss, J. C., '22. Manual of Instruction for Use with the Army Alpha Intelligence Tests. Third edition.

Kelley, T. C., '21. The Reliability of Test Scores.

Proctor, W. M., '21. The Use of Psychological Tests in the Educational and Vocational Guidance of High School Pupils. *Journal of Educational Research Monographs*, No. 1.

Yerkes, R. M., and Yoakum, '20. Army Mental Tests.

SOME FACTORS OF AESTHETIC JUDGMENT

Introduction

By Mrs. Anne Ross Reymert

The following experiments were undertaken between January and June, 1916, with a view of getting some light upon the various factors entering into aesthetic judgment.

After various preliminary tests upon miscellaneous observers, under different conditions, had been undertaken, the material, method and procedure for the main tests were determined.

In order to get at least a rough outlook over some of the main factors in aesthetic judgment, it was attempted to isolate, or to get these factors to *stand out,* one by one, in each of a series of different aesthetic preference-arrangements, thereby giving a means of comparison of the different rôle the specific factor may play, and its interplay with other factors in simple and more complex aesthetic judgments.

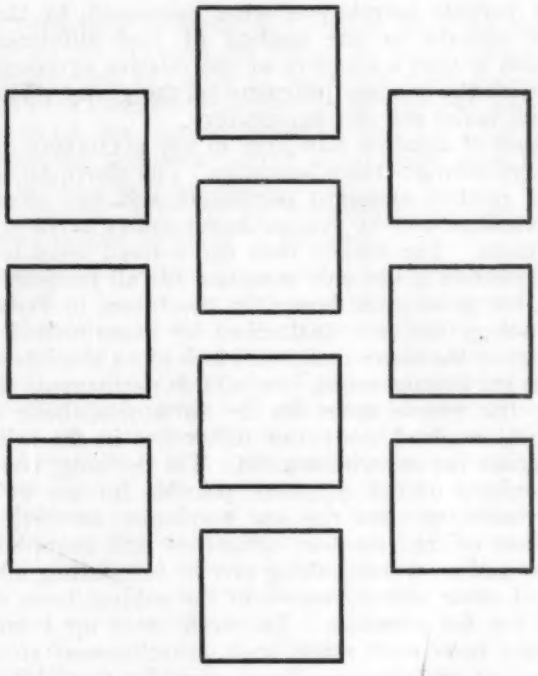
Twenty-two different series of aesthetic values were selected, each with the view of being:

1. a representative group series with reference to one specific aesthetic value, and
2. a comparable group in individual pictures (or card) value.

The observers consisted of two groups, one of 12 girls and the other of 12 boys, and were chosen at random among university students, mostly sophomores, juniors, and seniors, with two graduate students. All observers were without experimental training. The experimental conditions were kept uniform, and no tests were undertaken unless a subject reported himself in good condition.

The observers were experimented upon individually. The pictures were placed in a group arrangement as shown on inserted card. They were distributed within this plan so that no picture's relative position to any other picture might influence the observer's judgment, i.e., the scheme was to thoroughly mix up the pictures within the plan of arrangement.

This same arrangement, which afforded each picture an equal stimulus value, then was carefully maintained for all observers throughout all experiments. From Series S to Series EE, which involve the more complex stimuli, reasons



for the chief points upon which choice was based were requested.

Each observer's attention was carefully and explicitly directed to the one particular value to be judged in each series, and the instructions were kept uniform.

The observers were kept from communication with each other until tests were over.

The stimulus-cards or pictures were labelled with numbers on the reverse side where possible; otherwise they were lettered in a top-corner with a view of being unnoticeable and without meaning to the observers. The proportion tests, were labelled according to increase in width. To avoid influence of black and white effects in the proportion tests white outlines on black background were used.

Method

From the sum of the positions assigned to each individual picture by each group of subjects the average position or the average ranking of the picture was determined. And from

this the various correlations were calculated, by the method of least squares or the method of rank-differences. The correlation is thus a measure of the relative agreement of the judges with the average judgment of the group. The various additional tables are self-explanatory.

The lack of absolute standards in any preference judgments has always been a serious handicap. The above gives perhaps the best relative standard obtainable and has already been used advantageously by various investigators in the psychology of judgment. The results thus far derived seem to indicate that the method is not only adequate for all purposes of comparison, but gives great hopes for the future in various fields of psychology hitherto unattacked by experimental methods on account of the above-mentioned lack of an absolute standard.

To get ten homogeneous, comparable components that could make up the several series for the various aesthetic values to be judged, involved the prime difficulties in the selection of the materials for experimentation. The devising, constructing and organizing of the apparatus possible for use as a means of experimenting upon the *one particular* aesthetic element in isolation offered no less difficulties and required infinite exactness and most painstaking care in forestalling all possible sources of error and digression of the subject from the value exposed for his attention. To check these up a number of preliminary tests were made upon miscellaneous subjects.

Various art schools were drawn upon for available and possible material, suited to the testing out of the various aesthetic values, for which no apparatus was constructed. More had to be discarded than could be selected on account of the above mentioned prime difficulty.

The Various Experiments

The first three experiments consisted of the charcoal drawings group, Series A. B. and C. These particular drawings were mounted on a black screen in a similar general arrangement as mentioned above for all material used throughout these experiments.

The room for experimentation in this group was absolutely black, having no color effects in it whatever, and the only light that entered was the daylight coming from several small high windows. The screen was placed so that the light fell directly upon it at an angle of about 45 degrees. The observers were requested to seat themselves comfortably before the screen and contemplatively study these three series in

order, focusing their attention in Series A. upon the technique, i.e., excellency of drawing, in Series B. upon preference of pattern design, and in Series C. upon the most pleasing suggestiveness of a concrete object,—thus basing and noting their judgments upon one specific factor in turn for each series.

The following table may serve as a rough map over the substance of the problem, material, etc.:

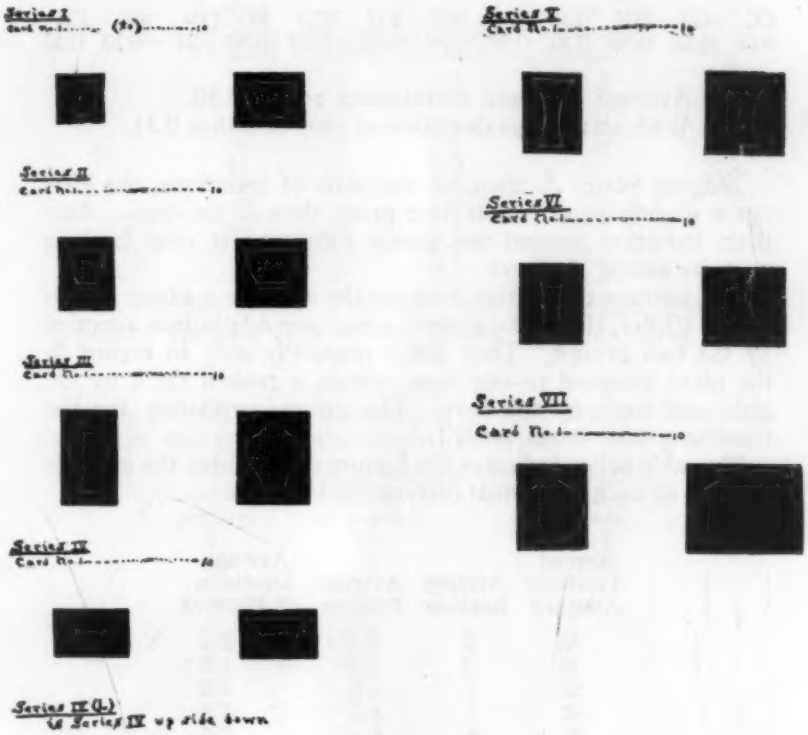
Tests: 22 Series, each of 10 comparable aesthetic values respectively; judged by same 12 subjects.

No.	Series	Groups	Kinds of Materials	Respective single factors as bases of judgment—i. e., factors tested
1.	A.	Charcoal Drawings	Drawings of similar degrees of merit of same study. (2 simple pieces of pottery) drawn from slightly different angles.	Technique
2.	B.		Drawings of pattern designs—each different.	Simple preference
3.	C.		Drawings of swirls—each different.	Suggestiveness of concrete object
Proportion tests:			Photostats—white outlines on black background mounted on white cards—all of equal size. Seven representative types: of non-geometric figures, consisting of varying relations between two curved lines and two straight lines.	
			Each series having height of its figures equal and equally graded increments in width from 1 to 10.	
			Series V and VI have the same figure but VI has a narrower proportion in width for card 1 and a wider proportion for card 10, and consequently larger increments in width serially than Series V.	

4.	I.	Proportion Tests	For proportion figures See p. 7 (a)	Width of entire figure increased.	Simple preference
5.	II.			Width of entire figure increased.	Simple preference
6.	III.			Width increased only at middle of figure.	Simple preference
7.	IV.			Width increased excepting at top.	Simple preference
8.	IV(L).			Width increased excepting at bottom.	Simple preference
9.	V.			Width increased only at top of figure.	Simple preference
10.	VI.			Width increased only at top of figure.	Simple preference
11.	VII.			Width of entire figure increased	Simple preference
12.	A & B	Same magazine cover illustrations		10 Arbor and Bird Day Magazine covers	Simple preference
13.	A & B (Color)			(Colored)	Preference based on color only
14.	A & B (Design			(Colored)	Preference based on design only

			Non-geometric	
15.	D.	Ceramic art designs mounted on cards of same size.	All have same general proportions and same five colors in different arrangements.	Simple preference
			Geometric	
16.	I.	Ceramic art designs mounted on cards of same size.	Designs with outstanding feature of complications of the lines—same five colors as in D.	Degree of intricacy of design
17.	S.	Illustrated Spring advertisements of similar size and mounted on cards of same size.	Non-colored. Word-matter removed.	Degree of suggestiveness of the idea of Spring
18.	AA.	Dutch Costumes.	Nine scenes representing Dutch costumes —colored.	Picturesque effect.
19.	BB.		Colored reproductions of paintings.	Preference with regard to desire as permanent possession in subject's room
20.	CC.	Post card reproductions.	Ten portraits of women. (Photogravures)	Most effective portrayal of feminine characteristic.
21.	DD.		Ten portraits of men. (Photogravures)	Most effective portrayal of masculine characteristic
22.	EE.		Ten religious pictures. (Photogravures)	Degree of impressiveness

PROPORTION FIGURES REDUCED ONE-HALF



EXPERIMENT I

Series A. Drawings of similar degrees of merit of the same study, two simple pieces of pottery, drawn from slightly different angles. The judgment was based upon the technique, or excellency of drawing.

Correlating in Series A. each individual girl judgment with the average of the girl group ranking the following correlations were found:

(Small letters refer to girls and capital letters to boys)

mf	me	tk	ec	ml	ad	ar	gl	mr	bt	ch	gs
0.30	0.14	0.43	0.75	0.65	0.55	0.48	0.37	0.71	0.68	0.38	0.49

Average of above correlations equals 0.49.
With an average deviation of plus or minus 0.14.

The same results for the boys:

CC	GJ	HH	LB	CS	NE	LH	WD	EO	DN	RG	EK
0.59	0.25	0.90	0.45	0.70	0.88	0.62	—0.30	0.90	0.24	—0.13	0.40

Average of above correlations equals 0.46.
With an average deviation of plus or minus 0.31.

Judging Series A. then, on the basis of technique, the girls agree slightly more within their group than do the boys. Also their variation around the group judgment is only half as great as among the boys.

The average correlation between the sexes as a group in this series, (0.49), shows the general agreement of position assigned by the two groups. They differ markedly only in regard to the place assigned to one item, which is ranked third by the girls and tenth by the boys. The group correlation for the remaining nine items is 0.71.

The table below indicates the factors determining the average ranking of each individual picture for the girls.

Sum of Positions Assigned	Average Ranking	Average Position	Average Deviation of Pictures
88	9	7.3	plus 2.3
71	5	5.9	or — 1.5
22	1	1.8	1.2
62	4	5.2	2.4
75	8	6.3	2.4
48	2	4.0	2.0
73	7	6.1	2.4
72	6	6.0	2.2
61	3	5.1	2.1
88	10	7.3	1.6

Pictures ranked number 1, 10 and 5 clearly show the smallest average deviations of position. Picture ranked number 1, owes its position through being ranked first by 9 of the 12 observers. Picture number 10, however, owes its position to a conflicting set of judgments, i.e., to spread votes. Similarly, picture ranked number 5 also owes its position to spread votes. From this the conclusion follows that clearly the greatest degree of agreement falls for the girls on the picture most preferred by both girls and boys.

Similarly, the following table for the boys:

Sum of Positions Assigned	Average Ranking	Average Position	Average Deviation of Pictures
72	6	5.8	plus 2.2
78	7	6.5	or — 1.7
23	1	1.9	0.9
62	4	5.2	1.8
81	9	6.8	2.5
52	2	4.3	2.4
62	3	5.2	2.7
66	5	5.5	2.0
87	10	7.3	2.5
77	8	6.6	1.9

Thus, for the boys it is obvious that the greatest degree of agreement likewise falls on the picture most preferred.

EXPERIMENT II

Series B. Charcoal drawings of pattern design. Here the judgment was based upon simple preference.

Labels of pictures	Average ranking by girls as a group	Average ranking by boys as a group
1	10	10
2	3	1
3	9	8
4	5	3
5	2	2
6	8	7
7	1	5
8	4	6
9	6	4
10	7	9

The individual correlations of the girls with their average ranking as a group are as follows:

mf	me	tk	ec	ml	ad	ar	gl	mr	bt	ch	gs
0.77	0.43	0.76	0.71	0.78	0.78	0.84	0.59	0.82	0.71	0.76	0.77

The individual correlations of the boys with their average ranking as a group are as follows:

CC	GJ	HH	LB	CS	NE	LH	WD	EO	DN	RG	EK
0.89	0.55	0.52	0.76	0.89	0.13	0.98	0.85	0.44	—0.19	0.08	0.21

GIRLS:	BOYS:
Average correlation of individuals with group order is 0.73	Average correlation of individuals with group order is 0.51
Average Deviation equals ±0.08	Average Deviation equals ±0.31

Correlation between girls and boys as one group equals 0.77.

In judging these charcoal drawings of pattern design on

the basis of simple preference the following results then are obtained.

As a group the girls agree very highly and show little individual variation around the group judgment.

The boys correlate only moderately high in this experiment, —considerably less than the girls. However, their span of variability is four times greater.

Experiment III

Series C. Charcoal drawings of swirls. Judgment based on: most pleasing suggestiveness of a concrete object. Here then, a more complex factor was involved.

Labels of pictures	Average ranking by girls as a group	Average ranking by boys as a group
1	5	3
2	4	4
3	3	5
4	10	9
5	9	7
6	2	2
7	8	6
8	1	1
9	6	10
10	7	8

The individual correlations of the girls with their ranking as a group are as follows:

mf	me	tk	ec	ml	ad	ar	gl	mr	bt	ch	gs
0.75	0.72	0.70	0.08	0.75	0.49	0.88	0.59	0.81	—0.31	0.77	0.88

The individual correlations of the boys with their ranking as a group are as follows:

CC	GJ	HH	LB	CS	NE	LH	WD	EO	DN	RG	EK
0.71	0.72	0.72	0.82	0.62	0.41	0.62	0.38	0.41	0.62	—0.36	0.64

Girls:		Boys:	
Average correlation of individuals with group order equals	0.59	Average correlation of individuals with group order equals	0.53
Average Deviation	±0.25	Average Deviation	±0.21

Correlation coefficient between girls and boys as one group equals 0.79.

Here again as a group, the girls agree more than do the boys; however, here they show, contrary to the former experiment, a higher degree of individual variation.

The attitude of the observers to Series C. was distinctly contemplative and the response in judging it distinctly more weighed and considered than in the previous experiment. From the remarks of the observers about their mental setting it was

apparent that the imaginative reaction played an insignificant part compared with the aesthetic reaction. In the imaginative reaction *no* judging factor was involved, hence the response much more immediate. However, with the entrance of the aesthetic reaction a critical, selective element entered, which was readily apparent from the comments of the observers upon the material. The dual process involved first, from the conditions of the test, that the aesthetic reaction be preceded by imaginative reaction and checked up by recognition. The characteristic then, of the aesthetic reaction seems to be distinctly contemplativeness.

Comparing the qualities of the aesthetic reaction in the three experiments, we may designate the first, Series A., as both distinctly and distinctively *objective* aesthetic reaction, and Series B. and C. purely *affective;* B., however, being simpler, and having less of the critical, is therefore, to a lesser degree contemplative.

From the comparison the average of group correlations for all these three experiments:

	Girls	Boys	Both Sexes as One Group
Series A	0.49	0.46	0.49 (or 0.71 for
Series B	0.73	0.51	0.77 nine pictures)
Series C	0.59	0.53	0.79

it seems, in contrasting the affective (B. and C.) with the objective (A.), as if there were greater agreement in more affective aesthetic judgments, both within the sex groups and between the sexes as one group.

In the above experiments there seems to be an increase of agreement between the sexes, with increase of the complexity.

Taking the average correlation of the group correlations, for girls, boys, and the two sexes together, from the above three experiments:

Girls 0.60 Boys 0.50 Both Sexes as One Group 0.68

it appears, that in aesthetical judgments based on technique, simple preference, and pleasing suggestiveness of a concrete object, the girls will approach more closely the mixed-group standard than will the boys, presumably because their preferences are less widely spread, i.e., more closely concentrated.

Experiments IV to XI

Proportion Tests (for materials see outline above).

Judgment based on simple preference.

Series I to VII

Average rankings in all proportion tests of girls and boys as separate groups:

CHART X

Labels of Cards	Series I		Series II		Series III		Series IV		Series IV(L)		Series V		Series VI		Series VII	
	G	B	G	B	G	B	G	B	G	B	G	B	G	B	G	B
1.....	10	7	8	10	8	10	10	10	8	10	9	3	9	10	10	10
2.....	8	9	5	9	6	8	8	7	2	7	7	6	8	9	9	9
3.....	5	3	6	7	1	4	6	6	4	5	5	2	5	7	8	8
4.....	3	6	3	5	2	2	5	3	3	6	3	5	1	4	7	7
5.....	1	1	1	2	4	3	4	2	7	4	4	1	4	3	4	6
6.....	2	2	2	1	5	1	1	1	1	1	1	4	3	1	2	1
7.....	4	4	4	3	3	5	3	4	5	2	2	8	2	5	1	2
8.....	6	5	7	6	7	6	2	5	6	3	8	10	6	2	3	3
9.....	7	8	9	4	9	7	7	8	9	8	6	7	7	6	5	5
10.....	9	10	10	8	10	9	9	9	10	9	10	9	10	8	6	4

CHART Y

Labels of Cards	Average Rankings	Average Positions	Average Deviations
1.....	10	8.8	Plus 1.3 or—
2.....	8	7.3	1.4
3.....	5	5.1	1.5
4.....	4	4.1	1.6
5.....	2	3.1	1.5
6.....	1	1.8	0.9
7.....	3	3.8	1.3
8.....	6	5.3	1.5
9.....	7	6.9	1.2
10.....	9	8.7	1.2

The above "Chart X" gives indications in reference to the *factors determining* the ranking of the cards, in all proportion tests.

Taking both sexes together, the greatest degree of agreement of actual votes falls both on the most preferred and on the least preferred card. Between these two extremes themselves, however, by far the greatest degree of agreement by actual vote falls on the most preferred card. The other cards owe their position chiefly to distribution of spread votes, but these—as may be seen from the average rank and also the average position of the cards in Chart Y—distribute themselves in regular increasing succession above and below the most preferred card. However, in glancing over the actual agreement of votes of the girls and of the boys separately, it is apparent that the boys agree more decidedly on both their most as well as on their least preferred cards, while the girls agree more emphatically on their least preferred card than they do on their most preferred card. The most preferred card of the girls owes its position more to spread votes than to actual agreement. Their least preferred card owes its position to their most emphatic agreement of choice. Considering the medium choices of each sex, the greatest actual choice among the boys falls on card labelled 3, and that of the girls on card labelled 8.

As *one group* of people then, the girls and boys together agree in the liked and disliked by actual votes, i.e., in the extremes.

As separate groups, the girls agree by actual vote only on their least preferred extreme; the boys by actual vote on both extremes, and on these choices only.

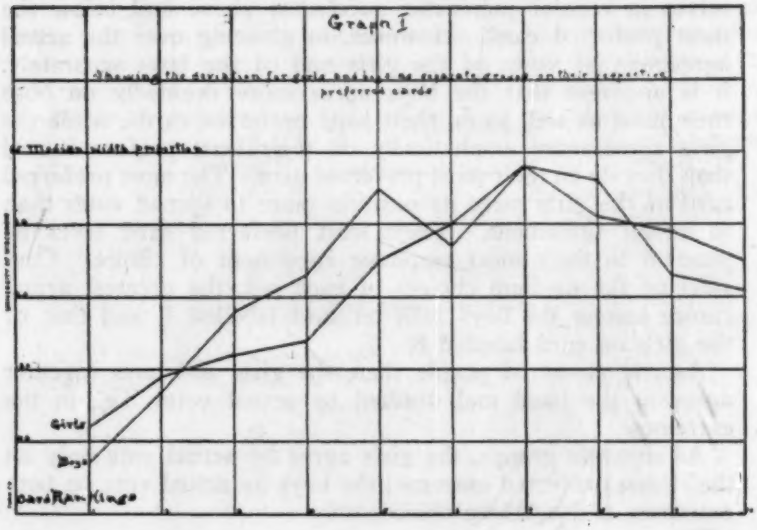
That is, together the sexes converge on likes but not on dislikes. Separately, the girls agree only on dislikes among themselves; the boys both in likes and dislikes.

CHART Z

Grouping the average ranking of the sexes together:

Card Labels	Sum Average Position	A. D.	Average Ranking	Average Ranking	Sum Average Position	A. D.
	Girls			Boys		
1.....	72	+17	9	10	70	+15
2.....	53	— 2	7	8	64	+ 9
3.....	40	—15	5	6	42	—13
4.....	27	—28	3	4	38	—17
5.....	29	—26	4	2	22	—33
6.....	17	—38	1	1	12	—43
7.....	24	—31	2	3	33	—22
8.....	45	—10	6	5	40	— 5
9.....	59	+ 4	8	7	53	— 2
10.....	74	+19	10	9	66	+11

The most strikingly outstanding feature is the fact of great relative general average agreement between the girls and boys successively, reading from card labelled 1 downward. But starting from the most preferred card (labelled 6) there is a successive distribution toward extremes in both sexes, yet with a gradual successive divergence above and below toward the extremes.



As is readily seen, from Chart Z, the middle choices of the two sexes, i.e., choices five and six, fall on cards labelled three and eight, showing clearly that their medium judgments fall half-way between the most preferred and the least preferred card of each sex, i.e., cards labelled one and ten. Then it may be concluded that, as two separate sex-groups, the sexes agree on the most preferred and the medium preferred cards. Also, Chart Z shows that the distribution of the deviations follows successively the distributions of the choices; the most preferred card, (number six), having the greatest deviation; the medium preferred a lesser deviation; then cards number two and number nine having the least deviation, and cards one and ten again having a medium deviation. Also, Graph I clearly indicates the similar direction of the choices of the two sexes.

The fact that there is one *definite, single, most-preferred* proportion figure, the medium one, and the further fact, that again there is a midway point of agreement between this median figure and the two extremes of proportion differences suggest: since the cards were labelled successively from one to ten in accordance to increase in width, thus giving a regular rate of proportions between two curved and two straight lines, it may be inferred that there probably is a medium preferred non-geometric type of proportion between two curved and two straight lines. The boys were emphatic in their first choice by actual average agreement, and also placed this agreement into the wider half of the proportions. When it came to choosing between the narrowest and the widest extreme proportions as the least preferred one the girls and boys were equally emphatic in their choice, *but* so in diametrically opposite directions.

May there then be a medium type of curved and straight line combination figure calling out a uniform aesthetic appeal?

May there be a "golden" proportion of this type analogous to the golden proportion of straight lines, or to the line of Hogarth? The kind of aesthetic response it calls forth seems to be very similar. Are the basic elements in this response of physiological origin?

To give a general view over the correlative results in all proportion tests the following table is inserted. It substantiates in the main the above inferences.

INDIVIDUAL CORRELATIONS IN ALL PROPORTION TESTS

	mf	me	tk	ec	ml	ad	ar	gl	mr	bt	ch	gs
Series I												
(G.)	0.53	0.85	0.85	0.25	0.53	—0.27	0.61	0.04	0.26	—0.04	0.02	0.54
(B.)	0.85	0.16	0.79	0.50	0.33	0.37	—0.02	0.28	0.50	0.14	0.38	—0.32
Series II												
(G.)	0.88	0.84	0.52	0.73	0.76	0.28	0.87	—0.32	0.32	0.95	0.33	0.41
(B.)	0.71	0.43	0.49	0.62	0.90	—0.07	0.75	0.35	0.42	0.66	—0.36	0.48
Series III												
(G.)	0.59	0.79	0.33	0.75	—0.05	0.84	0.77	—0.44	0.16	0.81	0.43	0.45
(B.)	0.62	0.42	0.76	0.81	0.53	0.81	0.58	0.49	0.87	0.13	0.13	0.47
Series IV												
(G.)	0.38	0.84	0.10	0.77	0.30	0.68	—0.22	—0.27	0.37	0.96	0.65	0.98
(B.)	0.54	0.37	0.90	0.48	0.60	0.71	0.84	0.26	0.20	0.82	0.53	0.60
Series IV(L)												
(G.)	0.09	0.50	0.76	0.59	0.64	—0.45	0.65	0.47	0.84	0.65	0.78	0.25
(B.)	0.89	0.66	0.58	0.38	0.94	0.37	0.01	0.50	0.52	0.72	—0.37	0.18
Series V												
(G.)	0.71	0.83	0.84	—0.03	0.49	0.10	0.05	—0.07	0.07	0.81	0.09	0.13
(B.)	0.82	0.66	0.62	0.10	0.41	—0.55	0.85	0.72	0.67	—0.19	0.19	0.36
Series VI												
(G.)	—0.05	0.87	0.15	0.15	0.94	0.05	0.03	—0.05	0.05	0.92	0.12	—0.03
(B.)	0.84	—0.36	0.82	0.78	0.28	0.81	0.84	0.61	0.94	0.07	0.37	0.55
Series VII												
(G.)	0.66	0.99	—0.20	0.98	—0.59	0.66	0.89	0.66	0.95	0.60	0.66	0.76
(B.)	0.70	0.81	0.28	0.90	0.89	0.77	0.14	0.78	0.88	0.70	—0.41	0.71
	CC	GJ	HH	LB	CS	NE	LH	WD	EO	DN	RG	EK

Series	Average Correlation	Average Deviation	Correlation between Girls and Boys as one group
I. (G).....	0.35	±0.30	0.84
(B).....	0.33	0.24	
II. (G).....	0.55	0.29	0.65
(B).....	0.45	0.25	
III. (G).....	0.45	0.31	0.73
(B).....	0.55	0.19	
IV. (G).....	0.46	0.35	0.88
(B).....	0.57	0.17	
IV.(L) (G).....	0.48	0.26	0.59
(B).....	0.45	0.28	
V. (G).....	0.34	0.34	0.33
(B).....	0.39	0.34	
VI. (G).....	0.26	0.32	0.70
(B).....	0.55	0.30	
VII. (G).....	0.59	0.32	0.94
(B).....	0.60	0.29	

The median correlation for all judgments of the girls is 0.44±0.31.
The medium correlation for all judgments of the boys is 0.49±0.26.
The median correlation between boys and girls, as one group, is 0.71

Omitting, however, Series V (which involved the finest increments of width increase, and hence, on account of the fine discrimination required, the choices may have been determined largely by chance), we get a correlation between the sexes of 0.76.

The correlation in both of the sex groups is very low, somewhat lower for the girls than for the boys, and the deviation is almost as great as the correlation, the boys showing a somewhat lesser variability.

The span of the individual correlations for the girls is —0.59 to 0.99, and for the boys is —0.55 to 0.94, showing a similar range of great individual variability for both sexes.

The average deviations on the cards in Series V, as compared with those for all the cards in all proportion tests, are more than twice as great.

The average correlation between the sexes as one group in Series V is the lowest for all twenty-two experiments.

SERIES V

Labels of Cards	Girls: Card preference or ranking	Girls: Average position of cards	Girls: Average Deviation	Boys: Card preference or ranking	Boys: Average position of cards	Boys: Average Deviation
1.....	9	6.6	Plus 3.8 or—	3	4.7	Plus 3.0 or—
2.....	7	5.8	3.0	6	5.0	2.0
3.....	5	4.9	2.4	2	4.5	1.4
4.....	3	4.9	1.1	5	4.8	0.9
5.....	4	4.9	1.3	1	3.9	2.3
6.....	1	3.9	1.8	4	4.8	1.6
7.....	2	4.8	1.8	8	6.6	1.7
8.....	8	6.4	2.8	10	7.6	1.9
9.....	6	4.8	2.4	7	6.3	2.8
10.....	10	7.4	3.0	9	6.9	3.0

Average correlation for the girls as a group = 0.34
Average deviation = ±0.34
Average correlation for the boys as a group = 0.39
Average deviation = ±0.34
Average correlation between the sexes as one group = 0.33

In Series IV and IV (L) the same figure was used, in the first instance with the widened portion at the bottom, in the second, it was exposed upside down. The lesser correlation between the sex groups in the second case is perhaps accounted for by the fact that the increase of width of the figure may have been more in a direct line of vision in the first case, and that we put meaning into characters from the top downward, hence notice the increase in width more attentively in Series IV.

There is also a sex difference in these two series:

	Girls as a Group	Boys as a Group	Girls and Boys as a Group
Series IV—Average Correlation.....	0.46	0.57	0.88
Average Deviation.....	±0.35	±0.17	
Series IV(L)Average Correlation.....	0.48	0.45	0.59
Average Deviation....	±0.26	±0.28	

The more obvious increase in width in the first exposure seems to have affected the girls negatively, but the boys positively.

Roughly, there is in these tests the same general variability as in the previous charcoal group.

EXPERIMENT XII-XIV

Series	Judgment based on—
A & B	Simple preference
A & B—(Color)	Preference based on color.
A & B—(Design)	Preference based on design

For this group of experiments, illustrative, colored Arbor and Bird Day magazine covers were used and so arranged for exposure that especially the color values of each magazine cover had an equal chance of appeal.

From the average rankings in these three experiments, it is of prime interest to find the possible part color and design, respectively, have played in the first selection of the magazine covers, in which all factors had free play.

AVERAGE RANKINGS

Labels of Pictures	A & B Simple Preference		A & B Color		A & B Design	
	(G)	(B)	(G)	(B)	(G)	(B)
1.....	10	9	9	9	8	6
2.....	8	10	10	10	7	10
3.....	5	4	4	6	5	8
4.....	7	1	7	2	4	1
5.....	1	3	1	1	3	4
6.....	6	7	6	7	6	9
7.....	9	6	8	8	10	5
8.....	2	8	5	4	2	3
9.....	4	5	3	5	1	2
10.....	3	2	2	3	9	7

It is found that the correlation between *simple preference* and *color* is:

for the girls, as a group 0.89
for the boys, as a group 0.82

which goes to show that color in the general judgment has played a conspicuous part, much more so than design, as the correlation between *simple preference* and *design* is:

for the girls as a group 0.61 (if last picture omitted) correlation = 0.77
for the boys as a group 0.45 (if last picture omitted) correlation = 0.46

This substantiates the psychological primacy of *color to form* in the aesthetical development of the individual and the rôle it plays in aesthetic judgment.

Also the girls correlate higher, both in color as well as in design, with their general simple preference judgment, than do the boys.

The following table gives a clear oversight over the correlations of all three series independently:

The individual correlations of the girls with their ranking as a group:

Series:	mf	me	tk	ec	ml	ad	ar	gl	mr	bt	ch	gs
A. & B.	0.59	0.64	0.88	0.25	0.53	0.68	0.87	0.76	0.41	0.50	0.66	0.26
A. & B. (Color)	0.55	0.66	0.27	0.28	0.40	0.68	0.48	0.81	0.82	0.81	0.81	0.73
A. & B. (Design)	0.75	0.90	0.43	0.58	0.92	0.49	0.81	0.64	0.67	0.13	0.79	0.31

The individual correlations of the boys with their ranking as a group:

Series:	CC	GJ	HH	LB	CS	NE	LH	WD	EO	DN	RG	EK
A. & B.	0.89	0.53	0.67	0.13	0.78	0.20	0.45	0.21	0.40	0.66	0.58	0.33
A. & B. (Color)	0.62	0.54	0.07	0.05	0.54	0.04	0.27	0.47	0.50	0.59	0.60	—0.19
A. & B. (Design)	0.24	0.22	—0.01	0.79	—0.03	0.52	0.68	0.47	0.13	0.04	0.75	—0.35

Series		Average Correlations	Correlation between sexes as one group
A & B	G.....	0.59±0.16	0.43 (0.52 for 9 pictures)
Simple preference	B.....	0.49±0.20	
A & B	G.....	0.60±0.16	0.78
(Color)	B.....	0.34±0.25	
A & B	G.....	0.62±0.19	0.56
(Design)	B.....	0.29±0.29	

As is seen here, the girls correlate much higher within themselves than do the boys. The variations in the correlations of the girls and boys as separate groups, in these series, may suggest that girls have more of a general standard in such aesthetic judgments in which color and design are involved than the boys have. Obviously the sexes agree best on color preference.

Experiments XV—XVI.

Series D: Judgments based on design—
Series I: Judgments based on intricacy of design—
The materal here consisted of Spanish ceramic art designs, essentially decorative, with a similar basic motif and having the same five colors. The design in the first series were non-geometric.

AVERAGE RANKINGS

Series D.

Individual correlations of the girls with their average ranking as a group:

mf	me	tk	ec	ml	ad	ar	gl	mr	bt	ch	gs
0.77	0.50	0.77	0.82	0.71	0.75	0.76	0.67	0.70	0.78	0.22	0.77

Individual correlations of the boys with their average ranking as a group:

CC	GJ	HH	LB	CS	NE	LH	WD	EO	DN	DN	RG
0.83	0.59	0.59	0.78	0.88	0.41	0.40	0.82	0.38	0.14	0.75	0.47

Series I.

Individual correlations of the girls with their average ranking as a group:

mf	me	tk	ec	ml	ad	ar	gl	mr	bt	ch	gs
0.98	0.93	0.05	0.54	0.82	0.64	0.90	0.83	0.76	0.83	0.33	0.75

Individual correlations of the boys with their average ranking as a group:

CC	GJ	HH	LB	CS	NE	LH	WD	EO	DN	RG	EK
0.82	—0.07	0.26	0.98	0.87	0.95	0.92	0.89	0.33	0.48	0.09	0.64

AVERAGE RANKINGS

Labels of Pictures	Series D		Series I	
	Girls	Boys	Girls	Boys
1.....	2	4	10	9
2.....	10	10	7	10
3.....	5	7	5	8
4.....	6	8	9	7
5.....	1	1	4	6
6.....	3	2	6	5
7.....	9	5	8	4
8.....	4	3	3	2
9.....	7	9	1	3
10.....	8	6	2	1

It is found that between these aesthetic preference judgments of design as such, and judgments based on intricacy of design, the girls correlate negatively (—0.05) and the boys very low, (0.25), which fact may indicate that as soon as an intellectual element is carried into the aesthetic judgment the preference may change almost entirely,—away from a directly impressive, aesthetic response.

Series		Average correlations of girls and boys as separate groups	Average Deviation	Correlation between girls and boys as one group
D	G.....	0.69	plus 0.11 or—	0.77
	B.....	0.59	0.19	
I	G.....	0.70	0.20	0.70
	B.....	0.60	0.32	

Correlation between girls as a group for both series: —0.05.
Correlation between boys as a group for both series: 0.25.

The greater variability in Series I as compared with Series D also makes us surmise that the entrance of the intellectual interferes with the full reaction of the aesthetic. The relatively equal correlations of the sex groups for both series indicate that there are great similar agreements within the individuals of the groups—for an entirely different aesthetic ranging—given the different conditioned judgment bases of Series D and Series I.

There is also great agreement between the sexes in this respect.

TO BE CONTINUED IN JUNE NUMBER

MINOR STUDIES IN THE PSYCHOLOGY OF ADVERTISING

From the Psychological Laboratory of Indiana University

V. The Development of Art-Forms in Magazine Advertising

By Harry Dexter Kitson[1]

It is evident to even the casual observer that art plays a much greater part in advertising than formerly. Some conception of this fact may be secured from figures reported in an earlier number of this series,[2] showing that the number of illustrations increased on the average from 26 in a hundred to 85 in a hundred during the quarter century from 1895 to 1919.

There should, however, be further differentiation in describing this development; for simultaneously with the increase in number of pictures has gone great advancement in kindred respects—photography, color process, etc. Furthermore, all apart from the mere matter of quantity has gone great advance in quality, until the advertisement of today is to that of thirty years ago as a pastel is to a chromo. This advance has come about by reason of the development in technique and the evolution of various art forms. It was to analyze these out and express their development quantitatively that this investigation was undertaken.

The data for the investigation were secured from the files of two representative weekly magazines—*Literary Digest* and *Harper's Weekly*. (The latter discontinued publication in 1916 and so *Collier's* was substituted after this date.) All the advertisements were examined (with the exception of those on the cover pages which had been removed in binding), for the years 1895, 1900, 1905, 1910, 1915, 1920. The illustrations were classified as follows: pen and ink, wash drawing, photograph, pencil, charcoal and combination (any one of these combined with another or with color).

[1] With the assistance of R. D. Bruce and G. M. Dykes.

[2] H. D. Kitson, Amount and Rate of Increase in the Use of Illustrations, *Jour. of App. Psychol.*, March, 1921, p. 12.

TABLE I

Showing the number and per cent of illustrations in the Literary Digest and Harper's Weekly from 1895 to 1920, Classified according to form of art used

	Pencil		Pen Ink		Wash		Photo		Combination		Char-Coal		Misc.		Total
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	
1895															
Lit. D.	0	0	1128	96	16	1.	37	3.	2	0	0	0	0	0	1182
H. W.	16	1.5	878	85.4	59	5.7	65	6.3	5	.48	3	.29	1	.09	1027
Total	16	.7	2006	90.7	75	3.3	101	4.6	7	.3	3	.1	1	.04	2210
1900															
Lit. D.	0	0	350	80.	80	17.	12	3.	3	0	0	0	0	0	445
H. W.	17	1.6	580	57.1	176	17.3	151	14.8	84	8.2	2	.19	4	.39	1014
Total	17	1.2	930	63.8	256	17.2	163	11.1	87	5.9	2	.1	4	.2	1459
1905															
Lit. D.	1	0	821	68.	177	15.	121	10.	79	7.0	1	0	0	0	1200
H. W.	56	8.4	310	46.9	66	16.	130	19.7	80	12.1	3	.45	15	2.2	660
Total	57	3.0	1131	60.7	243	12.3	251	12.9	159	8.6	4	.2	15	.8	1860
1910															
Lit. D.	6	.1	1353	41.3	947	29.	832	24.5	139	4.2	0	0	0	0	3277
H. W.	42	6.4	325	49.7	115	17.6	78	11.9	90	13.6	1	.15	2	.3	653
Total	48	1.2	1678	42.7	1062	27.	910	23.1	229	5.7	1	.02	2	.05	3930
1915															
Lit. D.	10	.2	1330	36.1	1136	27.4	642	15.2	1086	25.7	1	.0	4	0	4209
H. W.	4	1.4	151	54.3	48	17.3	28	10.	45	16.2	2	.7	0	0	278
Total	14	.3	1481	33.0	1184	26.1	670	14.9	1131	25.2	3	.06	4	.09	4487
1920															
Lit. D.	31	.6	1102	21.8	1402	27.8	1021	20.2	1483	29.3	3	0	5	0	5047
Colliers	77	3.6	448	21.	765	31.4	204	9.7	566	26.3	3	.14	13	.6	2076
Total	108	2.1	1550	21.5	2167	30.3	1225	17.1	2049	28.7	6	.08	18	.2	7123
Grand Total	260	1.2	8776	41.6	4987	23.6	3321	11.	3662	17.2	19	.09	44	.2	21069

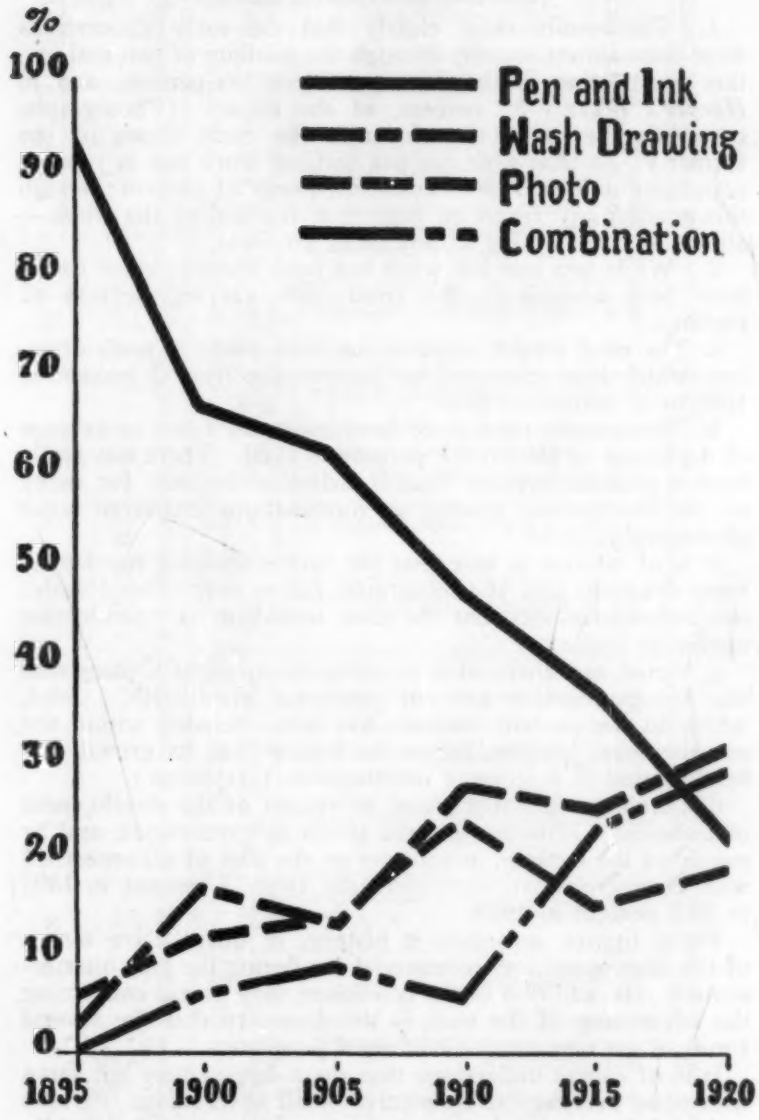


FIG. 1.—Showing per cent of illustrations in advertisements, 1895-1920, classified according to form of art employed

Results and Conclusions

1. The results show clearly that the early illustrations were done almost entirely through the medium of pen and ink, this constituting in the *Literary Digest* 96 percent, and in *Harper's Weekly* 85 percent, of the whole. (Photographs constitute the only other form in the early issues of the former.) As time went on, pen and ink work lost in relative popularity until in 1920 it constituted only 21 percent; though this amount still forms an important fraction of the whole—almost as large as that of any other art-form.

2. While pen and ink work has been waning, other forms have been coming to the front with varying degrees of rapidity:

a. The most notable advance has been made in wash drawings which have increased on the average from 3 percent in 1895 to 30 percent in 1920.

b. Photographs used alone have increased from an average of 4 percent in 1895 to 17 percent in 1920. There has really been a greater increase than is indicated by this, for many of the illustrations classed as combinations employed some photography.

It is of interest to note that the curves showing number of wash drawings and of photographs follow each other closely; this reflects the fact that the same technique of reproduction applies to both.

c. Pencil and charcoal have occupied so slight a place that the figures therefor are not presented graphically. Color, which in the present instance has been included within the combinations, involves factors so unique that its growth has been treated in a separate investigation (see below).

d. Combination illustrations, by reason of the development of technical skill in treating the forms just mentioned, and by reason of the increase in audacity on the part of all concerned with commercial art, have increased from .3 percent in 1895 to 28.7 percent in 1920.

These figures constitute a history, in quantitative terms, of the development of commercial art during the past quarter-century. In addition to the conditions they reveal concerning the advertising of the past, is the discovery that the several forms of art now enjoy about equal popularity.

It is of course understood that these figures may not serve with equal accuracy as descriptive of all advertising. Figures gathered from a study of other magazines might be slightly different, and those gathered from newspapers might differ more widely still. Still the facts here portrayed show the

TABLE II

Showing number and per cent of colored advertisements in four magazines, 1905-1920. (Files for certain years incomplete.)

	American			Colliers			Ladies Home Journal			Country Life			Totals		
	Full Page	Color Used	Per Cent	Full Page	Color Used	Per Cent	Full Page	Color Used	Per Cent	Full Page	Color Used	Per Cent	Full Page	Color Used	Per Cent
1905.....	0	0	0	x	x	x	0	0	0	901	10	1.1	901	10	1.1
1906.....	0	0	0	x	x	x	0	0	0	907	15	1.6	907	15	1.6
1907.....	0	0	0	563	44	7.2	0	0	0	815	14	1.6	1378	58	4.2
1908.....	0	0	0	513	34	5.6	35	1	2.7	660	11	1.6	1268	46	3.6
1909.....	0	0	0	439	11	2.4	54	6	10	846	25	2.8	1339	42	3.1
1910.....	0	0	0	644	18	2.7	26	5	16.1	774	7	.8	1444	30	2.0
1911.....	0	0	0	798	23	2.8	62	3	4.6	753	17	2.2	1613	43	2.6
1912.....	0	0	0	862	36	4.0	74	6	7.5	1076	28	2.5	2012	60	2.9
1913.....	189	49	27	980	17	1.7	80	3	3.6	901	28	3.0	2150	97	4.5
1914.....	134	61	45	838	28	3.2	64	4	5.8	983	19	1.9	2019	152	7.5
1915.....	110	44	44	956	78	7.5	41	24	37	787	10	1.2	1294	156	12
1916.....	204	61	30	1168	90	7.2	106	54	8.8	826	36	3.9	2304	241	10.4
1917.....	299	60	20	1224	64	4.9	177	138	43.8	768	36	4.4	2468	298	12
1918.....	414	72	17	915	50	5.1	192	208	52	558	34	5.7	2079	364	17.5
1919.....	715	122	16	707	42	5.6	317	391	55.2	671	39	5.4	2410	594	24.6
1920.....	959	221	24	652	93	12.4	956	481	50	382	97	20.2	2949	892	20.2

trends significantly, and offer a factual instead of a speculative basis for statements regarding the evolution of art forms in advertising.

VI. Color in Magazine Advertising [3]

In order to determine the extent to which color has been used in magazine advertising during the great expansion of advertising during recent years, a statistical study was made as follows:

In four representative magazines—*American, Collier's, Ladies' Home Journal* and *Country Life*—count was made of the percent of colored advertisements among full page advertisements (color being used only with full page advertisements), exclusive of cover pages.

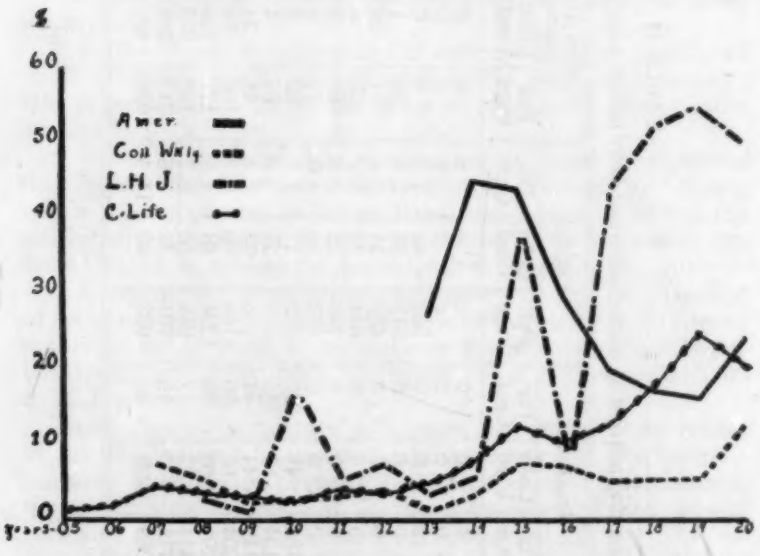


Fig. 1.—Showing per cent of colored advertisements in four magazines, 1905-1920

[3] With the assistance of N. Hottel, E. Sample, M. E. Smith, and R. Stegemeier.

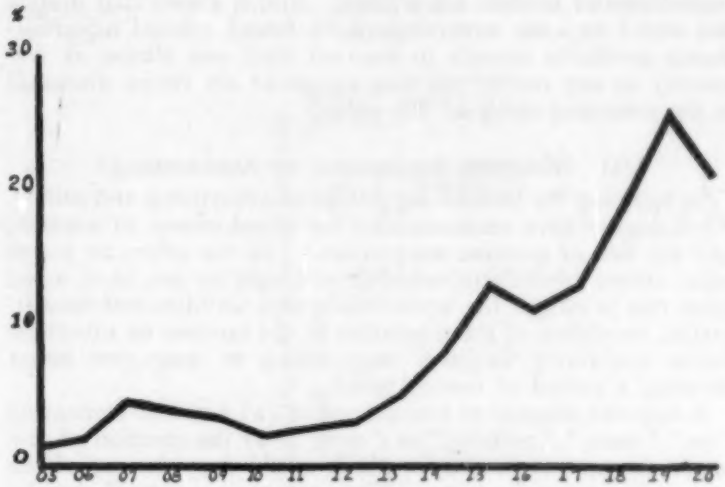


FIG. 2.—Showing per cent of total colored advertisements in four magazines, 1905-1920

RESULTS AND CONCLUSIONS

Results (as shown in Table I and Figures 1 and 2) show that *Collier's* and *New Country Life* were the first of these periodicals to use colored advertisements; the former as early as 1905, the latter in 1907.

After the *Ladies' Home Journal* and *American* entered the field of color in 1908 and 1913 respectively, there was a slight continuous increase until the period of the opening of the World War, when the rise became considerable. The slight decrease of 1916 may reflect a difficulty in securing imported pigments. Recovery was rapid, however, and under the stimulation of post-war business expansion the average amount rose to 24 percent. The decrease of 1920 doubtless reflects the decline of business in general.

Figure 1 shows marked differences between the amounts of color in the several periodicals. *Collier's*, even at its greatest height, has not exceeded a ratio of 12 percent, while the *Ladies' Home Journal* rose to 55 percent in 1919. (Examination of a copy of this magazine for December 1921 showed 89 percent of the full page advertisements to be colored).

The great increase here depicted, most of which has occurred within the past ten years, shows very dramatically the marvellous growth that has occurred in respect to this one of the

ingredients of modern advertising. And it shows that despite the added expense advertisers have found colored advertisements profitable enough to warrant their use almost as frequently as any one of the less expensive art forms discussed in the preceding study of this series.

VII Negative Suggestion in Advertising[4]

In applying the laws of suggestion to advertising and selling, psychologists have recommended the abandonment of negative and the use of positive suggestions.[5] In the effort to see to what extent advertisers, whether wittingly or not, have acted upon this principle, the writer undertook an historical investigation, consisting of the tabulation of the number of advertisements containing negative suggestions, in magazine issues covering a period of twenty years.

A negative suggestion was defined as (a) a phrase containing "no," "none," "nothing," or "not;" (b) the mention of any act on the part of the buyer which would deny the use of the commodity advertised, e.g., "Do not wear yourself out by sweeping;" (c) the mention by word or through illustration of any article competing with the one being advertised.

The advertisements examined were taken from three representative magazines: *Harper's Weekly, Literary Digest* and *Collier's*. Six months' issues were examined at intervals of five years, beginning with 1900. All the advertisements, exclusive of those on cover pages, were observed, and count was made of the number containing negative suggestions.

---

[4] With the assistance of J. H. Bales, M. M. Shirley, and R. C. Thomas.

[5] See, for example, W. D. Scott, Influencing Men in Business, N. Y., Ronald Press, 1st ed. 1911; 2d ed. 1915, p. 159; and H. D. Kitson, The Mind of the Buyer, N. Y., Macmillan, 1921, pp. 159-163.

TABLE I

Showing number of Advertisements Containing Negative Suggestions during period 1900-1920

	Harper's Weekly			Colliers			Literary Digest			Total		
	No. of Advs.	No. with N. S.	% with N. S.	No. of Advs.	No. with N. S.	% with N. S.	No. of Advs.	No. with N. S.	% with N. S.	No. of Advs.	No. with N. S.	% with N. S.
1900.....	1183	241	20.4				404	38	9.3	1587	279	17.6
1905.....	980	108	11.0	2196**	143	6.53	1275	58	4.5	4451	309	6.9
1910.....	640	91	14.2	2388	190	8.0	2019	130	6.4	5047	411	8.1
1915.....	239	16	6.7	1565	65	4.16	1989	107	5.4	3793	188	4.9
1920.....	***			1779	58	3.25	2543*	90	3.5	4322	148	3.4

*1919, **1903, ***Discontinued 1916.

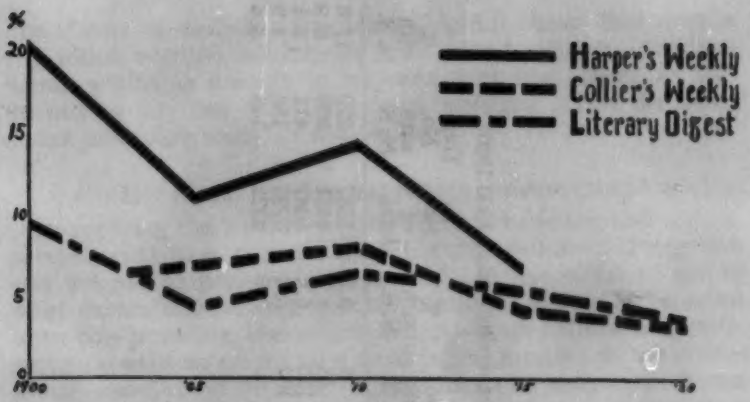


FIG. 1.—Showing per cent of advertisements containing negative suggestions in three magazines, 1900-1920

RESULTS AND CONCLUSIONS

Table I and Figure 1 show that of the advertisements investigated, 17.6 percent on the average contained negative suggestion in 1900. The number has now decreased, however, until there are only about 3.4 percent, or one-fifth as many.

The decline appears in all the periodicals examined, with fluctuations at approximately the same period.

The slight increase in 1910 may be due to the fact that advertising was taking on great vigor about that time and that in their zeal for achieving impressiveness advertisers were tempted to use quantity regardless of the value of their suggestions.

Of the three varieties of negative suggestion mentioned above, that of mentioning in a derogatory way a substituting or competing article was most common.

These results, constituting as they do, a record of the practices of advertisers, give basis for condemning the use of negative suggestions, all apart from theoretical grounds. In so far as historical evidence is indicative of the financial success of a mode of appeal, they indicate that advertisers have found it increasingly profitable to reduce the number of negative suggestions, until today they occur on the average only about 3 times in a hundred.

A SOCIAL RELATIONS TEST

By M. J. Ream, Bureau of Personnel Research, Carnegie Institute of Technology

There are, according to Professor Thorndike,[1] three types of intelligence—conceptual intelligence, mechanical intelligence, and social intelligence. Most of our present day intelligence tests tap the first mentioned type of intelligence but they make no pretense of measuring the last type, which functions in the social relations of life. This latter field is receiving increasing attention in experimental work. The Social Relations test here presented, attempts to throw light on this problem. It is intended to show the range and type of social interests. It is based on the assumption that the man who has a fund of such information, who has mixed with many social groups, is socially inclined and prefers to associate with people rather than spend his time at home or with books. Salesmen by reputation belong to this "social type." The test was accordingly given to two hundred salesmen. It has proved of some value in predicting the success of new salesmen.

The test consists of fifty items taken from about twenty lines of social diversion. The following are illustrations:

Socially acceptable items:

Games: Which of the following games requires chairs?
London Bridge..... Flying Dutchman..... Three Deep..... Going to Jerusalem.....

Lodges: In what organization is eleven o'clock of special significance?
Elks..... Odd Fellows..... Masons..... Knights of Columbus.....

Familiar songs: In the song, what follows the phrase, "I'm coming back to you, my Hula Lou"?
"So wait for me"..... "To make you mine"..... "In my canoe"..... "Beside the sea"..... "To old Hawaii" .....

Church songs: In the song, what follows the phrase, "Blest be the tie that binds"?
"Us in thy Kingdom, Lord"..... "My faith on Calvary" ..... "Loved ones of kindred minds"..... "Our hearts in Christian love".....

[1] Thorndike, E. L. Intelligence and its uses. Harper, 140: 227-35. Jan., '20.

Politics: What is a caucus?
A national political convention..... An official county election ..... A meeting of politicians within one party..... A secret political meeting in violation of law.....
Etiquette: What is "French leave"?
A dance..... Very few odds and ends left over..... Permission easily obtained..... Showing very polite manners ..... Slipping away without notice.....
Parliamentary practice: How may the maker of a motion withdraw it after it has been put before the house?
By a two-thirds vote..... By a majority vote..... By the consent of the chairman..... By the consent of the second .....
Slang: "Fussing" is a slang expression for what?
Fudging..... Nursing a sprain..... Playing the curb market..... Having a date.....

Sport items:
Baseball: What is the nickname of the Chicago Nationals?
Cardinals..... Braves..... White Sox..... Cubs.....
Golf: Which of the following golf clubs has a wooden head?
Brassy..... Niblick..... Mashie..... Cleek.....
Prize fighting: What kind of blow is a haymaker?
A hook..... An uppercut..... A broadside..... A jab.....
Horse racing: What is the name of the greatest living race horse?
Night Mist III..... Man O' War..... Dan Patch..... Sir Barton.....
Football: How many points are scored by a touchback?
0..... 2..... 3..... 6.....
Bowling: What is the score called when all ten pins are knocked down with two balls of one inning?
A strike..... A little slam..... A spare..... A break.....
General: What kind of race is a Derby?
Trotting..... Pacing..... Running..... Hurdling.....

Questionable items:
Dice: How many spots on dice make a "Little Joe"?
3..... 4..... 7..... 11.....
Liquor: Which of the following wines is the strongest?
Claret..... Champagne..... Sherry..... Burgundy..... Bordeaux.....
Cards: What beats a flush?
Fours..... A straight..... Three of a kind..... Two pairs .....
Theatre: What is the term applied to short, lively chorus girls?
Kittens..... Ponies..... Baby dolls..... Footlight dodgers .....
Dancing: Which of the following is best for jazz dancing?
Fox-trot..... Paul Jones..... Two-step..... Waltz..... March.....
Billiards: What is the name of the shot in billiards in which the cue ball is struck above the center?
Side shot..... Force shot..... Follow shot..... Draw shot .....

Roulette: What is usually the highest number on a roulette wheel?
21..... 36..... 40..... 55.....
Stock market: What does "cover on stock" mean?
Take it off the market..... Cash in..... Buy more of the same..... Put up money.....
Slang: What does the expression, "three fingers," stand for?
A glass of whiskey..... A glass of beer..... A pickpocket ..... A weak hand shake.....
General: What kind of news does the Polize Gazette chiefly contain?
Police court news..... Sport news..... Society news..... News of police organizations..... Abstracts of new laws .....

It will be noticed that the test contains three general classes of items, the socially acceptable, the sport, and the questionable items. While these classes of items appear to be mixed, they are so arranged that each group may be separately scored. The first half of the odd numbered items are socially acceptable, the second half of the odd items are sport and all the even numbered items are questionable. Relative standing on the three classes of items is very illuminating concerning a given individual's interests and habits.

The test follows the usual multiple choice principle of most information tests. The distractors have been carefully chosen to eliminate successful guessing as far as possible. In scoring, each correct answer is given one point. The test is untimed in order to eliminate the large speed factor operating in intelligence tests. Two alternate forms have been provided with fifty items in each form. The test in its original form contained one hundred items. It was divided into two forms on the basis of the difficulty of the items, which was determined by the number of times each item was missed. The resulting two forms showed correlations, on forty-eight cases, of .97 and .96 with the total score.

The Social Relations test was not intended to be an intelligence test, although it resembles the information type of standard intelligence tests. To determine to what extent the Social Relations test differed from the usual intelligence tests, correlations were computed with a special omnibus test adapted from the Army Alpha by the Bureau of Personnel Research.[2] The correlations on forty-eight cases were:

[2] An account of this test has been published. Bingham, W. V. Some group tests. *Psychol. Bull.*, February 1920, p. 57.

Social Relations test:	r
Score on socially acceptable items with Spiral omnibus intelligence test	.71
Score on sport items with Spiral omnibus intelligence test	.53
Score on questionable items with Spiral omnibus intelligence test	.38
Total score with Spiral omnibus intelligence test	.60

An intelligence factor is undoubtedly present but for inter-test correlations, these are not high.

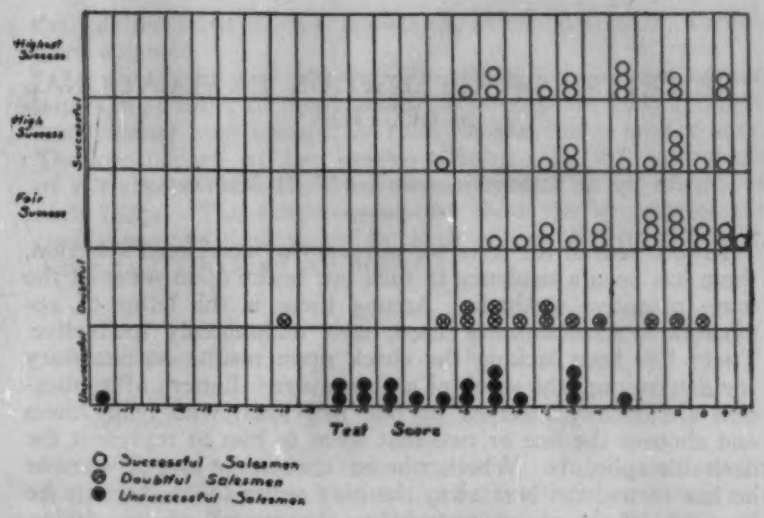
Other correlations were:

	r
Total score, Social Relations test with Education	.07
Total score, Social Relations test with hours spent in mechanical work	.09
Total score, Social Relations test with age	—.20

In considering the predictive value of the test for success in sales work, a detailed analysis of the separate items was made. A number of items gave statistical evidence of being significant though the total score on the test gave no such evidence. Older men are more successful salesmen while young men as a rule have higher total scores on the test. But when only the significant items of the test are considered in the scoring, discrimination appears. This is shown in the accompanying chart of insurance salesmen. Success in selling is determined only by amount of sales and number of cases sold while attending a three months sales school.

The test seems to have more negative than positive value for selecting insurance salesmen. In other words, ignorance of social diversions and lack of social contacts is unfavorable to success, but familiarity with many social diversions does not necessarily insure success in selling.

Aside from the test's predictive value, it is a valuable aid in throwing immediate light on an applicant's interests and his activities outside of working hours.



Legend for Chart

Fig. 1.—Predictive value of Social Relations test for success in selling insurance

X—Composite score on significant items of Social Relations test

Y—Success in selling insurance

THE LETTER OF APPLICATION IN VOCATIONAL SELECTION

By A. T. Poffenberger and V. H. Vartanian

In our search for tests for purposes of vocational selection, there has been a tendency to turn our backs upon some of the more primitive methods. Among these is the letter of application. As commonly used, it is undoubtedly ineffective. There has been lacking the check upon results so necessary for determining the value of any measure. Letters of application are usually received by one individual who reads them and chooses the one or two that seem to him to represent the desirable aplicants. Whether he has chosen the best or whether he has turned the best away, he may never know, because he has only the persons retained as the record of his choice. Perhaps his own judgment is not to be relied upon. His reaction to letters may be influenced by foolish prejudices either temporary or permanent. He might judge differently if he examined the letters again a week later; or the choice might be different if he were to ask the cooperation of ten of his business associates in evaluating the letters, and were to accept the results of their combined judgment. In spite, however, of its shortcomings as commonly used, the letter of application is still considered a part of the routine of the employment process by many institutions and individuals.

Now a letter of application, if really written by the applicant, might show many characteristics that are important and otherwise difficult to measure, such as ability to express oneself clearly and logically, neatness, good taste, seriousness of purpose and care for details. If checked up by some means as to accuracy of statement, it might conceivably serve as a good measure of honesty and integrity.

This report is intended to show to what extent the letter of application might be of service in vocational selection. One other study of similar nature will be briefly reviewed. Walton[1] studied the judgments, by fifty business and professional men, of twenty-five letters of application for a position as bookkeeper. Although there was no check upon

[1] An unpublished Master's thesis in the department of psychology of Columbia University.

the qualifications of the applicants, some interesting results were obtained.

1. *Consistency of the Judges.* Each judge arranged the letters in an order of merit twice, the interval between arrangements being long enough to rule out the factor of memory. The consistency of the judges is shown by the coefficients of correlation[2] between the first and second arrangement for each judge. The following figures show the average of the coefficients and their range for each of the four traits, intelligence, tact, reliability and neatness:

	Relation between first and second arrangement	Range of Coefficients
Intelligence	.537	.08-.72
Tact	.473	.18-.72
Reliability	.474	.20-.73
Neatness	.600	.14-.91

The figures show that the consistency of judgment varies with the trait; that individuals differ in their consistency in making such judgments; and that there are considerable differences among people in their ability to judge the trait itself from a letter.

2. *Agreement Among the Judges.* The judges agree more closely concerning the presence of some traits than of others when they are sought in letters of application. Those upon which there is most agreement are intelligence and tact, those upon which there is least agreement are reliability and neatness.

3. *Correlation Between Traits.* The coefficients of correlation between traits are given in the following figures:

Intelligence and Tact	.92
Intelligence and Reliability	.91
Intelligence and Neatness	.82
Tact and Reliability	.91
Tact and Neatness	.85
Reliability and Neatness	.84

They support what has frequently been reported, namely, that in such judgments it is not possible to react separately to particular traits, but rather the judge is governed by his general impression. Thorndike has aptly called this the "halo" effect.

4. *Judgment of Good and Poor Letters.* There is closer agreement upon the poor than upon the good letters for three of the traits judged, as shown by the following average devia-

[2] All coefficients of correlation presented in this report were obtained by the Spearman formula.

tions for the judgments in the case of the first and last letters:

A. D. for the 25th letter	4.02
A. D. for the first letter	5.84

If this were found to hold consistently, it would be advisable to select the best by elimination of the worst rather than by direct selection of the best.

The experiment to be reported here was performed several years ago. It differs from that of Walton in that it provided a check against which the judgments of the letters could be evaluated. Twenty-five students in the senior class of a training school for religious workers in New York City wrote letters of application for a position as religious worker in a certain section of the city. These letters just as received were given in turn to twelve members of the staff of the Union Theological Seminary, New York City, with the request that they arrange them in an order according to the degree to which they indicated general fitness for the position.

The measures which were to serve as checks against the letters were obtained as follows: Five teachers from the staff of the training school furnished three separate arrangements of twenty-five individuals according to the degree to which they possessed the three traits, general ability, intelligence and tact. These three traits were thought by the teachers to cover the qualities most needed for success in the religious work. In addition each member of the group of applicants arranged his twenty-four associates and himself in an order for each of these three traits. The conditions for such estimates were especially good, as the school is small and every teacher has very close association with his students; and the students themselves are more closely associated than are students in the ordinary college.[3]

We have then for purposes of comparison the following material:

a—Estimates of general fitness determined from the letters.
b—Estimates of general ability by the teachers.
c—Estimates of intelligence by the teachers.
d—Estimates of tact by the teachers.
e—Estimates of general ability by the associates.
f—Estimates of intelligence by the associates.
g—Estimates of tact by the associates.

[3] The average deviations of the positions assigned the individuals ranged from 1 to 5, where the total possible range is twenty-five. The average of these average deviations in the case of intelligence was 3.7. According to the formula presented by Berliner in the *J. of Appl. Psychol.*, 2, 1918, pp. 236ff, the average of all of the deviations for arrangements by chance would be 6.25.

1. *Correlation Between Traits.* When the individuals are judged directly by the teachers and associates, the traits are found to be related as follows:

General ability and intelligence	.95
General ability and tact	.82
Intelligence and tact	.86

These coefficients are about the same as those found by Walton where the letters were judged, and they indicate the tendency to judge from general impression, as previously stated. It would seem that, for practical purposes, a judgment for general fitness would be better than judgments for special traits, as such a task would be much less likely to confuse the judges.

2. *Relation Between the Estimates of Teachers and Associates.* The following coefficients show the relation between the average arrangement of the twenty-five individuals by the teachers and by the associates for the three traits:

	Relation between teachers' and associates' estimates
General ability	.90
Intelligence	.83
Tact	.59

The coefficients for general ability and intelligence are rather high, while that for tact is very much lower. This is, doubtless, due to the fact that the associates and teachers have different criteria of tact. There are no objective measures of this trait to form a common basis for judgment such as school grades would afford for intelligence or general ability.

3. *Estimates of the Letters.* The average deviations for the positions assigned to the individual letters ranged in size from 1.5 to 6.7 positions, where the total range of positions is twenty-five. This range is somewhat greater than in the case of the direct estimation of the individuals for intelligence (1-5). The average of all the deviations for all letters is 4.2 positions, where it would be 6.25 if the arrangements were chance arrangements. (See note above.) This is somewhat higher than the average of the deviations for the direct estimation of individuals for intelligence, namely 3.7. That is, there is a closer agreement in judging the individuals than in judging their letters.

As reported by Walton the agreement among the judges is closer for the poorest than for the best letter (A. D. of 1.5 as compared with A. D. 2.2), but as the difference is small and the relation is reversed for the next poorest and the next best, it has no particular significance.

4. *Relation Between Estimates of Letters for General Fitness and Direct Estimates of the Individuals.* The following figures show the correlations between the estimates of the letters and the estimates of the individuals for the three traits, general ability, intelligence, and tact by teachers and associates.

General ability (Combined estimates of teachers and associates) and letters	.50
General ability (Teachers' estimates) and letters	.56
General ability (Associates' estimates) and letters	.46
Intelligence (Combined estimates of teachers and associates) and letters	.44
Intelligence (Teachers' estimates) and letters	.58
Intelligence (Associates' estimates) and letters	.44
Tact (Combined estimates of teachers and associates) and letters	.22
Tact (Teachers' estimates) and letters	.20
Tact (Associates' estimates) and letters	.18
General ability, intelligence and tact (Combined estimates of teachers and associates) and letters	.45
General ability and intelligence (Combined estimates of teachers and associates) and letters	.48
General ability and intelligence (Teachers' estimates) and letters	.55
General ability and intelligence (Associates' estimates) and letters	.49

A consideration of these coefficients of correlation leads to the following observations:

Coefficients involving tact are the lowest. Walton, however, found rather close agreement among the judges for this trait. It has been noted earlier in this paper that the correlation between teachers' and associates' estimates of this trait was low. On the whole it seems unlikely that letters will provide a good measure of this trait.

Correlations between teachers' estimates and the letter estimates are higher than those between associates' estimates and the letter estimates. Various reasons for this difference suggest themselves, e.g., that the teachers are better judges, or that the two teacher groups (the judges of the letters being teachers) take a more similar attitude toward the task than teachers and associates would be expected to take.

Of the estimates by the teachers, that of general ability and intelligence both correlate fairly high with the average estimates of the letters, namely .56 and .58, in comparison with the usual test correlations. According to Thorndike, a general intelligence test could not be expected to correlate with success in college to a higher degree than about .60.

There is little relation between the estimation of tact from the letters and from the individuals themselves, whether the teachers or the associates are the judges.

If the average positions assigned for the different traits are combined and an order be determined therefrom, the correlation with the order for the letters is not raised above that for one trait alone. The same is true when combinations of two traits are correlated with the letter order.

5. *Relation Between Individual and Group Judgments.* The coefficients of correlation of individual arrangements of the letters with the average order for general ability obtained from the associates are as follows in order of size:

.18, .24, .26, .34, .34, .40, .40, .40, .46, .46, .48, .52

The average of these coefficients is .37. Now the coefficient of correlation of the *average order* for the letters (group judgment) with the average order of the individuals judged by their associates for general ability is .46. Thus the group judgment is somewhat better than the average of the individual judgments. Only two of the individuals show a higher correlation than that of the group. Therefore, although one might find a judge who would do better than the group judgment, the group judgment would be safer unless one had some means of knowing the good judges beforehand.

Now if the same comparisons are made with the teachers' arrangement instead of the associates' arrangement, the following coefficients are obtained, in order of size:

.24, .30, .34, .38, .40, .46, .49, .49, .53, 56, .57

The correlation for the average arrangement of the letters with the teachers' estimates is .56, as compared with the average of the individual coefficients .435.

6. *Consistency of the Judgments of the Letters.* Repeated arrangements of the letters by the same judges were not obtained, because the original letters of application were used and they had become soiled and worn with much handling. Since the conditions for judgment were very much the same as those in Walton's experiment we may assume that the consistency would be about the same.

7. *Basis for Estimation of the Letters.* There were numerous criteria for forming judgments of the letters, such as choice of paper, general style of the letter, handwriting, marginal spacing, line spacing, neatness, and mode of expression. There was also the actual content of the letter, such as the description of previous training, interest in the work, reasons for making application for the position, etc. No effort has been made to determine the importance of these various factors, although it would seem quite worth while to do so. The more mechanical factors such as those first listed could be

eliminated from the judgment, by presenting uniformly typewritten copies of the letters. The criteria would then be largely content and mode of expression. In the light of the usual requirement that letters of application be written in the applicant's own handwriting, it would be interesting to know the importance of these factors.

Conclusion. Individual differences in the ability to judge fitness from letters of application must be recognized. Some traits are much more clearly indicated in letters than others, while certain traits like tact can probably not be measured in letters. The combined estimates of a group of judges is more to be relied on than the judgment of one individual unless there is some means of picking the good judge. Some of the correlations here reported are as high as has been expected from the use of intelligence and other more specialized tests as measures of fitness for practical purposes.

The letter of application is not to be relied on even for a preliminary weeding out of candidates unless the ordinary safeguards against personal bias, etc., be used. But if treated and controlled as a test method it should be of service in vocational selection. It might be made to give a composite picture of neatness, intelligence, schooling, truthfulness, interest in details, ambition and many other of the character traits for which at present there are no adequate measures.

NEWS AND COMMENT

Notes Concerning Foreign Psychologists

The appearance has just been made of the first number of the Journal of the National Institute (British) of Industrial Psychology, recently noted in these columns. The aims of the Journal are editorially stated as follows:

"This Journal will describe in non-technical language the methods and results of applying scientific knowledge to the human aspects of industry and commerce.

"Its aim, therefore, forms part of the objects of the Institute which publishes it, namely, the promotion of the mental and bodily health and efficiency of every grade of worker. That excellence of work intimately and inevitably depends on the well-being, interest and suitability of the worker, is now generally recognized.

"The Journal will accordingly publish accounts of the research and propaganda work carried out in this wide field not only by the Institute, but by other similar bodies and by individual investigators both at home and abroad. It will contain abstracts of current periodicals and reports, and it will review the more important books on the subjects of industrial psychology and physiology. Contributions to the pages of the Journal are invited.

"The Journal will be issued gratis to every subscriber of the Institute, and it will appear quarterly or at such intervals as the material deemed worthy of publication allows."

This initial number contains brief, readable statements about the progress of several investigations being pushed in the fields of tin-box manufacture, chocolate packing, hiring of clerical workers. News is also given concerning vocational psychologizing in other countries.

In order for non-members of the Institute to secure the Journal, a minimum subscription price of one pound is set. The address of the editor is 329 High Holborn, London, W. C. 1.

There has been developing in Spain an application of psychology which has not been noted in these columns—the *Institut d'Orientació Professional* (Institute of Vocational Guidance) of Barcelona, (Sant Honorat No. 5). The Director is Sr. Ruiz Castella who is assisted by a staff consisting of a statistician, anthropometrist and a psychometrist, the latter being Dr. Mira I. Lopez.

The first three numbers of the *Anals*, the official organ of the Institute, just received, describe investigations relative to the measurement of individuals in relation to various occupations, after the usual patterns.

The latest number of the *Anals* of the Institute of Vocational Guidance at Barcelona announces that the *Institut de Psychologie* of the University of Paris has created a Section of Applied Psychology in charge of M. J. M. Lahy.

A publisher's (Barth, Leipzig) list of brochures containing studies in vocational psychology carried on chiefly under the direction of Drs. Stern and Lipmann includes reports of the application of psychology to street-car driving, typesetting, typewriting, telegraphy, aviation and hair-dressing.

American psychologists working in the field of business psychology are being honored by a visit from Dr. Y. Uyeno, Director of the Institute of Industrial Psychology of Tokyo Imperial University. Dr. Uyeno who was a student of Dr. Matsumoto (Yale) has been giving courses for some years in Psychology of Advertising, and more recently has been engaged in making application of psychological methods to industrial problems, chiefly with the end of increasing the efficiency of workers. He has translated Angell's Psychology into Japanese, and has written several books dealing with mental tests and their applications. On leave of absence this year, he is making a tour around the world, visiting laboratories in America and Europe. H. D. KITSON.

BOOK REVIEWS

ELIZABETH KEMPER ADAMS. *Women Professional Workers.* New York: The Macmillan Company, 1921, pp. 467.

This book, consisting of the fruits of a study made for the Women's Educational and Industrial Union, was "designed; first, for the thoughtful undergraduate who is trying to select her occupation, or, having selected it, trying to see its professional and social relations; second, for the young woman two or three years out of college or school who has drifted into a 'dead-end' occupation or from one occupation to another, and is now trying to get her bearings, and to take a longer view; third, for teachers and administrators in colleges and schools—deans, principals, vocational advisers, and so on—who are seeking a broader basis in fact and outlook in their dealings with young people; fourth, for employers who are increasingly turning to the colleges and professional schools for young men and young women workers; fifth, for men and women everywhere who are considering the scope and nature of the professions and the implications of the participation of women in the worlds of inquiry and affairs."

It takes in some sense the place of the earlier *Vocations for the Trained Woman* (published 1910) superseding the latter by reason of its greater wealth of information and the inclusion of the many new avenues opened to women by the stressful changes of the past decade.

The author would prefer not to treat the professional problems of women separately from those of men, and hopes that the day may come when it will be neither necessary nor profitable so to do. So far as possible, then, she tries to make the treatment of each profession interesting to workers of both sexes; giving scales of wages, kind of training required and illuminating facts about the nature of the work to be done.

Professions treated are Medicine, Law, Ministry, Food and Living Services, Civic and Government Services, Social Services, Personnel Services, Industrial, Commercial and Information Services, Art Services, Technical Services. Psychology is included in the last of these groups, and is accorded sage and generous treatment. The discussion might have been made somewhat more concrete by an investigation of the membership of the American Psychological Association, which in 1921 numbered among its 424 names those of 69 women—16 per cent.

The author notes the well-known fact that women have been regarding the teaching profession with increasing disfavor. In one of her tables showing occupations of graduates of three leading women's colleges, it appears that whereas in 1907 83 per cent entered teaching, in 1918 only 41 per cent entered the profession. The author deplores this fact and advocates that teaching and educational service in the new social setting constitutes the most important single profession for women.

Well conceived in its aim, and practical in its content, the book constitutes a real contribution to the vocational guidance of the growing number of young women who are enrolling in the professional schools of the country and are adding their talent and zeal to the upbuilding of the modern social structure.

A wisely selected and annotated bibliography adds greatly to the value of the book. H. D. K.

JAMES DREVER. *The Psychology of Industry.* New York: E. P. Dutton & Company, 1921, pp. 148.

BERNARD MUSCIO. *Lectures on Industrial Psychology.* New York: E. P. Dutton & Company, 1920, 2nd ed. rev., pp. 300.

CHARLES S. MYERS. *Mind and Work.* New York: Putnam, 1921, pp. 175.

The appearance of these three books within one twenty-four month period indicates a ripening interest in industrial psychology in the British Isles. They all treat the standard phases of industrial psychology—Vocational Fitness of Workers, Fatigue Study, Time and Motion Study—and all skillfully plead that industry be organized on a scientific basis with an eye to the welfare of the worker.

The slight differences between the three books consist chiefly of individualities in emphasis upon particular points: in the Muscio book arguments designed to meet the objections brought by labor against scientific management; in the Myers book a chapter on Systems of Wage payment; and in the Myers and Drever books the unfoldment of a dream of a central bureau of industrial psychology, similar to the National Institute in England and the Psychological Corporation in the United States. The Drever book contains also two chapters on the applications of psychology in advertising and selling and a series of six tests chosen from the Terman revision of the Binet series which, though accompanied by no norms, are alleged to "yield a result which is not greatly increased in accuracy by the addition of the others."

Upon reading the half dozen or more books dealing directly or indirectly with industrial psychology, the sophisticated psychologist-reader grows wishful that writers would introduce some variation from their unanimous citations of the Munsterberg street-car drivers, the Thompson bicycle ball inspectors, and the Taylor pig iron handlers. Written as they were, however, not primarily for the delectation of psychologists but for the enlightenment of intelligent laymen, these three books should not be expected to bring new contributions to science. The psychologist should be happy to have his case stated so carefully and attractively as it is in these books. And by reason of the experimental work insured by the fostering care of the National Institute of Industrial Psychology there is reason to look to the British Isles for an increasing number of contributions toward the too-scantily clothed body of industrial psychology.

MCKINNEY, JAMES, AND SIMONS, A. M. *Success through Vocational Guidance.* Chicago, American School, 1922, pp. 270.

The task of vocational guidance, which is usually regarded as one of measuring the individual at an early period of his life and thereupon designating him for some one vocation, is bound to give way with increasing enlightenment to a saner acknowledgement that the individual has a great number of potential vocational adaptations within him, but that he cannot make any of those that are worth while without training. The task of vocational guidance thus envisaged becomes that of directing the individual into channels of profitable training and education.

In the work of providing training for men and women seeking better vocational adjustment, which the American School of Correspondence has been carrying on for the past quarter-century, this

institution has been obliged perforce to abandon the theory that a human being is born cut to a vocational pattern. It has helped too many persons change the pattern through the instrumentality of education. It is from such a premise that this book was prepared by the Educational Director and the Director of Foreman Training of the School.

As befits a book designed for the practical, earnest clientele of the School, this volume formulates the problem in downright terms like these: "Every position should be a station on the road of life, not a terminus. Other things being equal, a job is desirable according to the number of roads leading from it to better jobs." "Most of us make many vocational choices. There is no final choice until life ends." Every occupation is thus regarded as a training place where the individual may prepare himself for further advance. Data are presented showing that with each degree of training goes a proportionate measure of advancement. Each occupation described is discussed in terms not only of the kind and degree of training required in it, but also in terms of the pathways which it opens to still higher forms of service.

The leading groups of occupations are treated with as great detail as is possible in the compass of 270 pages; and the reader who has access to the original sources will recognize that the facts are trustworthy.

The book should prove generally useful in the work of vocational guidance in public schools, especially in connection with courses in "Occupations." To the research worker it should prove provocative of investigations designed to supply the great stores of factual material which for the present such books as this are bound to want.

H. D. KITSON.

WILLIAM FREDERICK BOOK, Professor of Educational Psychology and Head of Department of Psychology, Indiana University. *Intelligence of High School Seniors.* The Macmillan Company, January, 1922, 391 pp.

Presents results of state wide mental survey of Indiana High Schools. Develops a practical technique for locating the brightest seniors graduating from high school with a view of providing better educational opportunities for them; determines and presents the range, general level and variations in intelligence of high school seniors going to college and colleges of different kinds, i.e. technical and liberal arts; the intelligence of seniors coming from different economic and occupational classes; the intelligence of those who had been retarded, accelerated, and regularly promoted in high school and the grades; the intelligence of the seniors selecting different high school courses and different occupational careers; the intelligence of seniors representing different communities and schools, etc. The book is divided into three parts: the first describes the method used in making the survey and sets forth why such mental surveys should be made; the second presents the results obtained in this State-wide mental survey. Section three discusses and interprets the results and points out their educational and social significance. It also shows how a mental survey may be used in helping to solve some of the more important practical problems which educators, social workers and progressive business men are called upon to solve, who seek to conserve, cultivate, or utilize the native mental capacities of children and

men. The book develops a method or technique which may be used in making mental surveys of communities and schools. Will be reviewed in later issue of this JOURNAL.

WILLIAM F. BOOK, Head of Department of Psychology, Indiana University. *An Efficient Method for Measuring the Results of Instruction in Colleges and Schools.* Bulletin Indiana University Extension, Vol. VI., No. 12, December 1921.

The report defines and describes an efficient method for measuring school accomplishment, one which will measure the success of a teacher or school and the scholastic achievement of pupils in terms of what may reasonably be expected from individuals with different grades of mental ability. The need for such a pedagogic method is illustrated by materials taken from the results of extensive mental surveys showing marked individual differences in the native mental endowment of the pupils belonging to the same school grades, in the same school system, differences between buildings, school systems taken as a whole, different types of communities, etc. But the chief part of the report is devoted to the results obtained in experimenting with such a method of measuring the scholastic attainment of 300 college students at Indiana University. A technique for making such measurements is worked out and described and the reactions to this more accurate and pedagogic method of reporting the school achievement of students is described and the conclusions drawn from this experiment set forth in detail.

"Personnel Research Agencies" is the title of bulletin No. 299 just issued by the United States Department of Labor, through its Bureau of Labor Statistics (No. 299). It is intended to serve as a guide to organized research in employment management, industrial relations, training, and working conditions.

About a year ago a conference was held in Washington, D. C., under the auspices of the Engineering Foundation and National Research Council to consider the practicability of bringing about cooperation among the many bodies conducting research relating to persons employed in industry and commerce, from management to unskilled labor. This conference, which was attended by 40 representatives of organizations of labor, manufacturers, employment managers, engineers, physicians, educators, economists, and social workers, requested the Bureau of Labor Statistics to undertake a survey of existing agencies whose activities include studies and investigations relating to the employee and his job. The above bulletin contains the results of this survey arranged in the form of a handbook for ready reference, it describes the research activities of about 300 bureaus, associations, foundations, laboratories, and university departments, which deal with the problems of an employment manager's office, vocational psychology, wage systems, cost of living and budgets, employee representation, training of managers, foremen and workmen, whether in educational institutions or in the factory, the relation of hours of labor, fatigue, lighting, ventilation, food, etc., to output and health of workers, occupational diseases and health hazards in the various industries, safety codes and appliances for the prevention of accidents, and the special problems connected with the employment of women and young persons, immigrants, colored workers, the handicapped or disabled, and the mentally deficient or unstable.

The industrial engineer, the employment manager, the trade-union official, the educational director in retail store or factory, the vocational counselor in the public schools, the industrial physician, and the social worker will all find information useful to them in this compendium.

NEW BOOKS RECEIVED

Until further notice books for review should be sent to W. F., Book Department of Psychology, Indiana University.

H. Addison Bruce. *Self-Development.* A Handbook for the Ambitious. Funk & Wagnalls Company, New York and London, 1921. 332 pp. Price, $1.50.

Grace Owen. *Nursery School Education.* E. P. Dutton & Co., New York. 176 pp.

Kenneth Richmond. *The Curriculum.* Constable & Company, London, 1919. 144 pp. Price, 5s.

Margaret McMillan. *The Nursery School.* E. P. Dutton & Company, New York; J. M. Dent & Sons, London and Toronto, 1919. 356 pp.

Janet Penrose Trevelyan (Preface by Mrs. Humphrey Ward). *Evening Play Centres for Children.* E. P. Dutton & Company 183 pp.

L. A. Kalbach and A. O. Neal. *Organization of State Departments of Education.* U. S. Bureau of Education, Bulletin, 1920, No. 46, Washington, D. C. 48 pp.

—— ——. *Survey of the Schools of Wilmington, Delaware.* U. S. Bureau of Education, Bulletin, 1921, No. 2, Washington, D. C. 191 pp.

The Journal of Educational Method, Vol. I, No. 3, November, 1921. The World Book Company, Greenwich, Conn. Price, $3.00 a year, 35c a copy. 128 pp.

Guy M. Whipple. *The National Intelligence Tests.* The Public School Publishing Company, Bloomington, Illinois. pp. 16-31.

Arnold Gesell. *Hemihypertrophy and Mental Defect.* (Reprinted from the Archives of Neurology and Psychiatry, Oct. 1921, Vol. VI). American Medical Association, Chicago. 24 pp.

Walter E. Fernald and George M. Kline. *Massachusetts Department of Mental Diseases,* Vol. IV, Nos. 3 and 4, October, 1920. 135 pp. Boston, Mass. (Papers in Honor of Dr. E. E. Southard's Decennial Anniversary of the Bullard Professorship of Neuropathology at Harvard University.)

Daniel Bell Leary. *Philosophy and Education: Part I.* University of Buffalo, 1921. University of Buffalo Studies, Vol. II, No. 2, October, 1921, pp. 45-89.

Sante Naccarati. *The Morphologic Aspect of Intelligence.* Columbia Contributions to Philosophy and Psychology, Vol. XXVII, No. 2, New York, 1921, 44 pp. (Report of Correlations between measurements of intelligence and measurement of extremities and of trunk reciprocally considered and compared.)

Raymond F. Crist. *Federal Citizenship Textbook.* (A course of instruction for use in the public schools by the candidate for citizenship.) U. S. Department of Labor, Washington, D. C., Part III, 104 pp.

Kentucky Educational Commission. *Public Education in Kentucky.* General Education Board, 61 Broadway, New York 1921, 213 pp.

GEORGE S. SNODDY and GEORGE E. HYDE. *Mental Survey of Utah Schools and Adaptation of the Army Beta Tests.* Bulletin of the University of Utah, Vol. 12, No. 6, Sept., 1921, pp. 5-28.

LILLIEN J. MARTIN. *The Work of a Mental Hygiene Clinic for Pre-School Age Children.* Publication No. 5, California Society for Mental Hygiene, San Francisco.

——— ———. *Pedagogical Hints from the Results of a Survey of a San Francisco Public School for Delinquent Boys.* Martin Mental Hygiene Publication No. 5, pp. 5-18.

WILLIAM F. BOOK. *Preliminary Report on the State-wide Mental Survey of High School Seniors.* Bulletin of the Extension Division, Indiana University, Bloomington, Indiana, pp. 31-67.

——— ———. *Variations in Mental Ability and its Distribution Among the School Population of an Indiana County.* Bulletin of the Extension Division, Vol. IV, No. 4, Indiana University, Bloomington, Indiana, 42 pp.

PERCIVAL HALL. *Education of the Deaf.* Department of the Interior, Bulletin No. 14, 1921, Washington, D. C., 16 pp.

——— ———. *Present Status of Music Instruction in Colleges and High Schools* 1919-20. Department of the Interior, Bulletin, No. 9, 1921, 53 pp. (Report of a study made under the direction of the United States Bureau of Education by a joint committee of the National Education Association, Music Teachers' National Association, and Music Supervisors' National Conference.)

GEORGE F. ZOOK and SAMUEL P. CAPEN. *Opportunities for Study at American Graduate Schools.* Department of the Interior, Bulletin, No. 6, 1921, 59 pp.

SAMUEL PAUL CAPEN. *Facilities for Foreign Students in American Colleges and Universities.* Department of the Interior, Bulletin, No. 39, 1920, 269 pp.

The Journal of Applied Psychology is anxious to purchase copies of the June, 1917 (Vol. 1 No. 2) issue.

Address

FLORENCE CHANDLER

Publisher, Clark University

Worcester, Mass.